





; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 3980  
; LENGTH: 765  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3980

Query Match 9.4%; Score 49; DB 4; Length 765;  
Best Local Similarity 45.2%; Pred. No. 0.14;  
Matches 219; Conservative 0; Mismatches 265; Indels 1; Gaps 1;  
QY 1 GGTGACAGCAAAACGACCGGTGGTGGCGCGCGCCCATCAGAACTGCCACTCCGGCG 60  
Db 64 GGGTCACTACCATCAGGTTGGCGCGCGCGCAACCGAGCAGGTTCGCGAGCACCTGG 123  
QY 61 GGGTACAGGTTCAGCAGGTGGAGTTGAGCAGCCAGCTCATCGCCGCTGCGCGGCA 120  
Db 124 CCTCGCTGGCGCGCTCAGAGGCTGGCGGTGGGCTTTGTCGAGCAGCAGCAACCGGCGCTCG 183  
QY 121 TGCCGAAACACCGGCGCAGGATCACGTGACAGCGGATGCGGCGCTTCGATCTCGACGG 180  
Db 184 AGACAGCCCGAGCGAGCAGGCGTTGCGCGTGGCGCGGAGCAGCAGCGCGCGC 243  
QY 181 TCGCGCGAGTGCATCTGTCGCGCGCTCCACACAGGGGAAATGAGTCTGGTGGA 240  
Db 244 CGGACCGGAGTCTGTCGCGCGCTCCGCGCGAGGCTCGATCTCTTCAGCAGGCGCA 303  
QY 241 GCGCAGCCCGCAGCGGAGCAGCTGTCAGGCTGGGCTCAGCGACTCTCGCGGTGG 300  
Db 304 CGGCGCGCGCTAGCGAGCGCTTCGAGATCGGCGTTCGCTCGCGCCTAGCGCAGGT 363  
QY 301 GAGCGGAAACGCGCAGCGCTCAGATCCCTGTGATCGATCGGCTCAGTCCGCTGCTCC 360  
Db 364 TCCAGGCAACCTGCGCGGAAACAGGCGTTGTCTCGAAC-ACCAGTTACGATTGGCG 422  
QY 361 CCCTTGGCTGGGAGTACGGTTACAGCAGCGGACACACGCGGCGGGGGGG 420  
Db 423 GCGAGGTGCTTTGCTCAACCGCGGATATCCACGCCATCCAGCAGGACACGCGCGCA 482  
QY 421 CGGCTTCAGCGATCCGCTCGATGACCGAGGCTTCGGGGTTCGGGGTTCGGCGAGTAC 480  
Db 483 TCGAGCTGTAGAGCGCGCGGAGCAGCGGCGGAGGCTGCTCTTGGCGCGCGGAGGT 542  
QY 481 CGTAC 485  
Db 543 CCGAC 547

RESULT 6  
US-09-252-991A-4015  
; Sequence 4015, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 4015  
; LENGTH: 2223  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (2110)  
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.  
US-09-252-991A-4015

Query Match 9.4%; Score 49; DB 4; Length 2223;  
Best Local Similarity 45.2%; Pred. No. 0.14;  
Matches 219; Conservative 0; Mismatches 265; Indels 1; Gaps 1;  
QY 1 GGTGACAGCAAAACGACCGGTGGTGGCGCGCGCCCATCAGAACTGCCACTCCGGCG 60  
Db 615 GGGTCACTACCATCAGGTTGGCGCGCGCGCAACCGAGCAGGTTCGCGAGCACCTCG 674  
QY 61 GGGTCAACAGGTTCGACAGGTGGAGTTGAGCAGCCAGCTCATCGCCGCTTGGCGGCA 120  
Db 675 CCTCGCTGGCGCGCTCAGAGGCTGGCGGTGGGCTTTGTCGAGCAGCAGCAACCGGCGCTCG 734  
QY 121 TGCCGAAACACCGGCGCAGGATCACGTGACAGCGGATGCGGCGCTTCGATCTCGACGG 180  
Db 735 AGAGCAGCCCGAGCGAGGACAGGCGTTGCGCGTGGCGCGGAGCAGCAGCGCGCGC 794  
QY 181 TCGCGCGAGTTCGATCTGTCGCGCGCTCCACACAGGGGAAATGAGTCTGGTGGA 240  
Db 795 CGGACCGGAGCTGCTGCCAGCGCTCGGCGAGGCTCGATCTCTTCAGCAGGCGCA 854  
QY 241 GCGCAGCCCGCAGCGGAGCAGCTGTCAGGCTGGGCTCAGCGACTCTCGCGGTGG 300  
Db 855 CGGCGCGCGCTAGCGCAGCGCTTCGAGATCGGCGTTCGCTCGCGCTCGCGCAGGT 914  
QY 301 GAGCGGAAACGCGCAGCGCTCAGATCCCTGTGATCGATCGGCTCAGTTCGCGGTCTCC 360  
Db 915 TCCAGGCAACGCTGCGCGGAGCAGCGCTTCTCTCGAAC-ACCAGTTACGATTGGCG 973  
QY 361 CCCTTGGCTGGGAGATAGCGGTTTCAGCAGCGGACACACGCGCGGCGGGGGGG 420  
Db 974 GCGAGGTGCTTTGCTCAACCGCGGATATCCAGCGCACCCAGCAGGACACGCGCGCA 1033  
QY 421 CGGCTTCAGCGATCCGCTCGATGACAGCGGCTCGGGTTCGGGGTTCGGCGAGATC 480  
Db 1034 TCGAGCTGTAGAGCGCGCGGAGCAGCGGCTGCTCTTCCGCGCGCGGAGGT 1093  
QY 481 CGTAC 485  
Db 1094 CCGAC 1098

RESULT 7  
US-09-252-991A-13914/c  
; Sequence 13914, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13914  
; LENGTH: 816  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13914

Query Match 9.3%; Score 48.6; DB 4; Length 816;  
Best Local Similarity 52.6%; Pred. No. 0.16;  
Matches 133; Conservative 0; Mismatches 114; Indels 6; Gaps 1;

QY 65 GACCAGGTCCACCAAGTGGGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGGATGCC 124  
DB 661 GACCAGGTGAGTGTGGTGTATCAGCAGCAGGACATGCCAGCGGTGCTTCAGGTC 602

QY 125 GAAACACCGGCGCAGGATCAGTGCAGCAGCGGATGCGCGCTCGATCTGACAGGTGG 184  
DB 601 CTTGAGCAGATCAGGATCTTCACTGACGGTGA-----CGTCAGCGCGGTGTCGG 548

QY 185 CCGCAGCTCGATCTCGTCGCGCGCTCCACACACAGCGGAACTGCTCGTGGCAGCG 244  
DB 547 TTGTCGGCGATCAGCAGTTGCGGCTCGCAGCGCGGATGCGCATCATCACCGCTG 488

QY 245 CAGCCCGCAGCGGACAGTCTGTCAGGCTGCGGTCAAGCACTCTCGCGGTCGGAGC 304  
DB 487 CCCTGCGCGCGCAGAGTTGCTGTGATAGGCTTTTCAGGCGCTTGGCGGCTCGGGAT 428

QY 305 GGAACCGCGCACG 317  
DB 427 GCCGACAGCTCG 415

RESULT 8  
US-09-252-991A-13437  
; Sequence 13437, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13437  
; LENGTH: 966  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13437

Query Match 9.3%; Score 48.6; DB 4; Length 966;  
Best Local Similarity 52.6%; Pred. No. 0.16;  
Matches 133; Conservative 0; Mismatches 114; Indels 6; Gaps 1;

QY 65 GACCAGGTCCACCAAGTGGGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGGATGCC 124  
DB 262 GACCAGGTGAGTGTGGTGTATCAGCAGCAGGACATGCCAGCGGTGCTTCAGGTC 321

QY 125 GAAACACCGGCGCAGGATCAGTGCAGCAGCGGATGCGCGCTCGATCTGACAGGTGG 184  
DB 322 CTTGAGCAGATCAGGATCTTCACTGACGGTGA-----CGTCAGCGCGGTGTCGG 375

QY 185 CCGCAGCTCGATCTCGTCGCGCGCTCCACACACAGCGGAACTGCTCGTGGCAGCG 244  
DB 376 TTGTCGGCGATCAGCAGTTGCGGCTCGCAGCGCGGCGCATGCGCATCATCACCGCTG 435

QY 245 CAGCCCGCAGCGGACAGTCTGTCAGGCTGCGGTCAAGCACTCTCGCGGTCGGAGC 304  
DB 436 CCCTGCGCGCGCAGAGTTGCTGTGATAGGCTTTTCAGGCGCTTGGCGGCTCGGGAT 495

QY 305 GGAACCGCGCACG 317  
DB 496 GCCGACAGCTCG 508

RESULT 9

US-09-252-991A-13705/c  
; Sequence 13705, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13705  
; LENGTH: 1668  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13705

Query Match 9.3%; Score 48.6; DB 4; Length 1668;  
Best Local Similarity 52.6%; Pred. No. 0.16;  
Matches 133; Conservative 0; Mismatches 114; Indels 6; Gaps 1;

QY 65 GACCAGGTCCACCAAGTGGGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGGATGCC 124  
DB 717 GACCAGGTGAGTGTGGTGTATCAGCAGCAGGACATGCCAGCGGTGCTTCAGGTC 658

QY 125 GAAACACCGGCGCAGGATCAGTGCAGCAGCGGATGCGCGCTCGATCTGACAGGTGG 184  
DB 657 CTTGAGCAGATCAGGATCTTCACTGACGGTGA-----CGTCAGCGCGGTGTCGG 604

QY 185 CCGCAGCTCGATCTCGTCGCGCGCTCCACACACAGCGGAACTGCTCGTGGCAGCG 244  
DB 603 TTGTCGGCGATCAGCAGTTGCGGCTCGCAGCGCGCATGCGCATCATCACCGCTG 544

QY 245 CAGCCCGCAGCGGACAGTCTGTCAGGCTGCGGTCAAGCACTCTCGCGGTCGGAGC 304  
DB 543 CCCTGCGCGCGCAGAGTTGCTGTGATAGGCTTTTCAGGCGCTTGGCGGCTCGGGAT 484

QY 305 GGAACCGCGCACG 317  
DB 483 GCCGACAGCTCG 471

RESULT 10  
US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 9.3%; Score 48.6; DB 3; Length 4403765;



Best Local Similarity 45.5%; Pred. No. 0.17;  
Matches 209; Conservative 0; Mismatches 249; Indels 1; Gaps 1;  
QY 57 GCGGGGTGACACAGTTCAGACAGTGGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCC 116  
Db 3925089 GCGGTGATCAATGCGCCACACAGCGCTGTGCGGGGCGCGTTGATCGGTGACGCGCC 3925148  
QY 117 GGCATCCGAAACACCGGCGCAGATCAGTGTGACAGCGCGATGCGGCGCTCGATTCG 176  
Db 3925149 AATGCGAGATCTCCGCGGGGGCGCGCGCGCTGCTGTGCGCAACGCGGG 3925208  
QY 177 ACGGTGCGCCCGAGCTCGATCTCGTTCGCGCCCGCTCCACACAGCGGAAACTGCTCGGT 236  
Db 3925209 CCGGGGCGAGCGGGGCGCGCGCCAGCGCGCGGCTGCGGGCGCGCGGATTGTG 3925268  
QY 237 GCGAGCGCAGCGCCAGCGCGGAGAGTGTGTCAGGTGCGGTGCGGTACGAGCTTCGCGGG 296  
Db 3925269 GGCACGCGCGCGCGCGCGCGCGCGCGCGCGGTGCGGGCGCGCGCGCGCG 3925328  
QY 297 TCGGAGCGGAAACGCGCAGCGCTCAGATCCCTGTGATGCGCATCGGCTCAGTCCGCGTC 356  
Db 3925329 CCGGTGGGTGGCTGTTCGGGTGGCGCGCGCGCGGTGTGATGGGCGCGGTGGCGGC 3925388  
QY 357 GTCCCGCTTGGCTGGAGAGATAGCGGTTCAGACAGCGCGCACACAGCGGGCGCGGC 416  
Db 3925389 ACCGCGCGCGCGCGCGCGCGCGGTGTTGATCTGCGCGCGCGCGCGCGCGGTGTC 3925448  
QY 417 GGGGCGGTTCAGCGCATCGCTCGATGACACAGCGGTGCGGGTTCGGGCGGTTCGCGCA 476  
Db 3925449 GGTGGGCGGTGGCGGACCGCGCGG - CCGGCGCGCGCGCGCGCGGTGCTGTTCGCGC 3925507  
QY 477 GATCGTACCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCG 515  
Db 3925508 CCGCGGTGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3925546

RESULT 11  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: PRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 9.3%; Score 48.6; DB 3; Length 4403765;  
Best Local Similarity 47.7%; Pred. No. 0.17;  
Matches 173; Conservative 0; Mismatches 189; Indels 1; Gaps 1;  
QY 154 GCGGATGCGCGCTCGATTCGAGGTGCGCGCGAGCTCGATCTCGTCCCGCGGTCCC 213  
Db 338941 CGCGCTGCTCGCCCGCATCAACAGTTCCTTGGCGAATACCGGCGCGCGCTGATCG 338782  
QY 214 ACACAGGGAACTGGTGGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 273  
Db 338781 GCAACGGACCAACG 338722

QY 274 TGGCGTCACGACCTCTCGCGGTGGAGCGGAAACGCGCACGGCTCAGATCCCTGTCA 333  
Db 338721 TGGCAACGCGCGCGCGCGGAGATCGGCGCGCGCGCGCTCAACGCGCGCGCGCGCA 338662  
QY 334 GTCCGATCGCTCAGTCCGCTGCTCCCTTGGCTTGGAGATAGCGGTTCACGACGA 393  
Db 338661 ACGCGCGCGCGCGCGCGCTCATCGGCAACGCGCGCGCGCGCGCGCGCGCGCGCGCA 338602  
QY 394 GCGCACCAACGCGCGCGCGCGCGGTTCAGCCGATCCGCTCGATGACACGCGCG 453  
Db 338601 GCACGGGACCG 338543  
QY 454 TCGCGGTGCGCGCGGTTCGCGAGATCCGTACGCGCGGACCGCTTCGCGCGCGCGCG 513  
Db 338542 GCGCGGTGCGCGGTTCGCGCGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 338483  
QY 514 GGG 516  
Db 338482 GGG 338480

RESULT 12  
US-09-252-991A-3898/c  
; Sequence 3898, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 3898  
; LENGTH: 606  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3898

Query Match 9.2%; Score 48; DB 4; Length 606;  
Best Local Similarity 48.2%; Pred. No. 0.21;  
Matches 135; Conservative 0; Mismatches 145; Indels 0; Gaps 0;  
QY 1 GGTGACGACGAAACGACCGGTGTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 347 GGTCACTACCATCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 288  
QY 61 GGTGACGACGATCGACAGGTGGAGTTGAGCAGCGAGCTCATCGCGCGCTGCGCGCGCA 120  
Db 287 CCTCGTGGCGCGTGGAGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGG 228  
QY 121 TCCGAAACACCGCGCGCGAGATCAGTGCAGACGCGCGATGCGCGCTTCGATTCGACGG 180  
Db 227 AGAGCAGCG 168  
QY 181 TCGCGCGCGAGCTCGATCTCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 167 CGGACCGAGCTCGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 108  
QY 241 GCGGAGCG 280  
Db 107 CG 68

RESULT 13  
US-09-417-704-2/c  
; Sequence 2, Application US/09417704  
; Patent No. 6444874

GENERAL INFORMATION:  
; APPLICANT: Duivick, John  
; APPLICANT: Guilliam, Jacob  
; TITLE OF INVENTION: A Hydroperoxide Lyase Gene from Maize and Methods of  
; FILE REFERENCE: Maize hydroperoxide lyase (HPL)  
; CURRENT APPLICATION NUMBER: US/09/417,704  
; CURRENT FILING DATE: 1999-10-13  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1835  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: (1) (115)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1025)..(1027)  
; OTHER INFORMATION: I-helix  
; FEATURE:  
; NAME/KEY: misc binding  
; LOCATION: (1457)..(1459)  
; OTHER INFORMATION: Heme-binding site  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: (1625)..(1835)  
; US-09-417-704-2

Query Match 9.0%; Score 47; DB 4; Length 1835;  
Best Local Similarity 50.2%; Pred. No. 0.34;  
Matches 116; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
QY 20 GTGGTGGCCCGCCGCGGATACGAACTGCGGCGGGGTGACCGAGTCCGACCG 79  
Db 1528 GTGCTGTAGCGCGGAAACAGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1469  
QY 80 GTGGGAGTTGAGCAGCAGCTATCCGCGCGCTGCGCGCGGCGGCGGCGGCGGCGG 139  
Db 1468 CTGGCGGCGCACTGTTGTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1409  
QY 140 GATCAGTGCAGCAGCGCGATGCGCGCTCGATCTGAGCGGTGCGCGCGGCGGCGGCTC 199  
Db 1408 GAAGAGGTGTCAGCAGCGCTGCGCGCTCGTCCGCGGAGGCGGTTCCGCGCACTC 1349  
QY 200 GTGCGCGCGCTCCACACGAGGAGAACTGGCTCGGTGCGGCGGCGGCGGCGGCGG 250  
Db 1348 CTGGGCGCGCTCGAACACCTCGGGGTCCGCGATCGCGCGGCGGCGGCGGCGGCGG 1298

RESULT 14  
US-09-103-840A-1  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
; US-09-103-840A-1

Query Match 9.0%; Score 47; DB 3; Length 4411529;  
Best Local Similarity 45.3%; Pred. No. 0.35;  
Matches 208; Conservative 0; Mismatches 250; Indels 1; Gaps 1;  
QY 57 GCGGGGTGACCAAGTCCAGGTCGACAGTGGAGTTGAGCAGCAGCTCATCGCGCCCTGCGCC 116  
Db 3931301 GCGGTGATCAATGCGCCACCCAGGCGCTGCTGGGGCGCCCTGTGATCGGTGACGGCGCC 3931360  
QY 117 GGCATGCCGAACACCCGGGCGAGATCACGTGCAGCAGCGCGATGCGCGCTCGATCTCG 176  
Db 3931361 AATCGACGACTCCCGCGCGGGCGCGGCGCGGCGCTGCTGTTCGCAACGCGCGG 3931420  
QY 177 ACAGTCCGCGCGAGCTCGATCTCGTCCCGCGGCTCCACACAGCGGGAATCGCTCGGT 236  
Db 3931421 GCCGGGCGAGCCGGGGCGCCCGGCGAGCGCGGGCTCGGGCGCCCGCGGATTTGG 3931480  
QY 237 GGCAGCGCGAGCCCGGCGAGCTCGTCCAGGCTGCGGTACAGGACTTTCGCGGG 296  
Db 3931481 GGCAGCG 3931540  
QY 297 TCGGAGCGGAAACGCGCGAGCGCTCAGATCCCTGTCAGTCGATCGGCTCAGTCCGCTC 356  
Db 3931541 GCCGGTGGGTGGCTGTTGGGGTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 3931600  
QY 357 GTCCCTTGGCTGGAGGATAGCGGTTACGACGAGCGCGCGCGCGCGCGCGCGCGCG 416  
Db 3931601 ACCG 3931660  
QY 417 GGGGGCGGTTTCAGCGGATCCGCTCGATGACAGCGGCTCGGGGCTCGGGCGGCTCG 476  
Db 3931661 GGTGGGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3931719  
QY 477 GATCCGTACCG 515  
Db 3931720 CGGCGGTGCGGCTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3931758

RESULT 15  
US-09-833-381-1616/c  
; Sequence 1616, Application US/09833381  
; Patent No. 6672186  
; GENERAL INFORMATION:  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs  
; FILE REFERENCE: 5800-119  
; CURRENT APPLICATION NUMBER: US/09/833,381  
; CURRENT FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 09/516,448  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 2050  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1616  
; LENGTH: 1977  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-833-381-1616

Query Match 9.0%; Score 46.8; DB 4; Length 1977;  
Best Local Similarity 48.0%; Pred. No. 0.37;  
Matches 165; Conservative 0; Mismatches 177; Indels 2; Gaps 1;  
QY 152 CACGGCGATCGCGGCTCGATCTCGACGGTTCGCGCGCGCGCGCGCGCGCGCGCGCTC 211  
Db 798 CCGCGGAAGCCAGCCCGGCTCTCGTGGGTAGCAGCGGCTGGGGCGCGCTCCAGCCT 739  
QY 212 CCACACGAGGAACTGGTTCGTTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 271  
Db 738 CCAGGCCAAGCGCGACCCACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 679  
QY 272 GCTGGCGTCAAGGACTCTCGCGGCGCGGAGCGGAAACGCGCGCGCTCAGATCCCTCT 331  
Db 678 GCGGATCCG 621

QY 332 CAGTCGCATCGCTCAGTGCCTGCTGCTCCCTTGGCTGGGAGGATAGCGTTCCAGAC 391  
Db 620 CGGTTAAAGGCGCAGGACTCTGCGGCCCGCCCGGTTGGCGGGGTGAGGGCAACGCTAA 561  
QY 392 CAGCGGCACCAACGCGCGCGCGCGCGGCTTCAAGCCGATCCGCTCGATGACCAAGCG 451  
Db 560 GGAACCCCTCAGCGCTCTCGGACTGCGGCGTGTCCCGCGGCCCAAGTTGAAACGCCCG 501  
QY 452 CTGCGGGGTGCGGGCGTGGCGAGATCCGTACCGCCCGGACC 495  
Db 500 CCAGAGCGCAGAGCGCGCTCGGGAACGTTTGAGACCGTTCC 457

RESULT 16  
US-09-252-991A-180  
; Sequence 180, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 180  
; LENGTH: 1461  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (489)  
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.

US-09-252-991A-180

Query Match 8.7%; Score 45.4; DB 4; Length 1461;  
Best Local Similarity 49.4%; Pred. No. 0.69;  
Matches 118; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 23 GTCCCGCGCGCGCCATCAGAACTGCCACTCGCGGGGTGACAGTGCACGAGTG 82  
Db 1197 GTCCACCTCGCGGCCAGCAGCCGCGGAGCAACGCCGCGAGAGGGCACTCGACCATGAT 1256  
QY 83 GGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGGATGCCGAAACACCGGCCAGGAT 142  
Db 1257 CCCAGTTGAGTTCGCCAGCGGGATTCTTCGCGAGCGCGAGCGCCAGGTACGCGC 1316  
QY 143 CACGTGACGACCGCGATGCGCGCTCGATCTCGAAGCGTGGCGCGCAGCTCGATCTCGTC 202  
Db 1317 CTGCGGCCATTCTGTCGAGGTGCGGACCATCGGAAACATCACCGCAGCGCGCTCCCC 1376  
QY 203 GCCCGGCTCCACACCAAGGGAACCTGGCTGGTGGCAGCGGAGCCCGCCAGCGGGACA 261  
Db 1377 GGCGGCGGNAAGCGCGGCACTGGTCTCGAGATCTGCGTCTGAGGGTCA 1435

RESULT 17  
US-09-252-991A-166/c  
; Sequence 166, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 166  
; LENGTH: 1560  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-166

Query Match 8.7%; Score 45.4; DB 4; Length 1560;  
Best Local Similarity 49.4%; Pred. No. 0.69;  
Matches 118; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 23 GTCCCGCGCGCGCCATCAGAACTGCCACTCGCGGGGTGACAGTGCACGAGTG 82  
Db 1188 GTCCACCTCGCGGCCAGCAGCGCGCGGAGCAACCGCGAGAGGGCACTCGACCATGAT 1129  
QY 83 GGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGGATGCCGAAACACCGGCCAGGAT 142  
Db 1128 CCCAGTTGAGTTCGCCAGCGGGATTCTTCGCGAGCGCGAGCGCCAGGTCAAGGAT 1069  
QY 143 CACGTGACGACCGCGATGCGCGCTCGATCTCGAAGCGTGGCGCGCAGCTCGATCTCGTC 202  
Db 1068 CTGCGGCCATTCTGTCGAGGTGCGGACCATCGGAAACATCACCGCAGCGCGCTCCCC 1009  
QY 203 GCCCGGCTCCACACCAAGGGAACCTGGCTGGTGGCAGCGGAGCCCGCCAGCGGGACA 261  
Db 1008 GGCGGCGGNAAGCGCGGCACTGGTCTCGAGGATCTGCGTCTGAGGGTCA 950

RESULT 18  
US-09-252-991A-175  
; Sequence 175, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 175  
; LENGTH: 2610  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-175

Query Match 8.7%; Score 45.4; DB 4; Length 2610;  
Best Local Similarity 49.4%; Pred. No. 0.69;  
Matches 118; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 23 GTCCCGCGCGCGCCATCAGAACTGCCACTCGCGGGGTGACAGGTGCACGAGTG 82  
Db 391 GTCCACCTCGCGGCCAGCAGCGCGGAGCAACCGCGAGAGGGCACTCGACCATGAT 450  
QY 83 GGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGGATGCCGAAACACCGGCCAGGAT 142  
Db 451 CCCAGTTGAGTTCGCCAGCGGGATTCTTCGCGAGCGCGAGCGCCAGGTCAAGGAT 510  
QY 143 CACGTGACGACCGCGATGCGCGCTCGATCTCGAAGCGTGGCGCGCAGCTCGATCTCGTC 202  
Db 511 CTGCGGCCATTCTGTCGAGGTGCGGACCATCGGAAACATCACCGCAGCGCGCTCCCC 570  
QY 203 GCCCGGCTCCACACCAAGGGAACCTGGCTGGTGGCAGCGGAGCCCGCCAGCGGGACA 261  
Db 571 GGCGGCGGNAAGCGCGGCACTGGTCTCGAGGATCTGCGTCTGAGGGTCA 629

RESULT 19  
US-09-773-816-1/c  
; Sequence 1, Application US/09773816  
; Patent No. 6340774  
; GENERAL INFORMATION:  
; APPLICANT: Stanford University  
; APPLICANT: Khosla, Chaitan  
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR  
; TITLE OF INVENTION: ANTAGONISTS  
; FILE REFERENCE: 28600-20210.00  
; CURRENT APPLICATION NUMBER: US/09/773,816  
; CURRENT FILING DATE: 2001-06-08  
; PRIOR APPLICATION NUMBER: US 60/243,458  
; PRIOR FILING DATE: 2000-10-25  
; PRIOR APPLICATION NUMBER: US 60/179,305  
; PRIOR FILING DATE: 2000-01-31  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 23673  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(23623)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-773-816-1

Query Match  
Best Local Similarity 8.7%; Score 45.2; DB 4; Length 23673;  
Matches 116; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
QY 56 CGCGGGGTGACAGGTGACAGGTGAGGAGTTGACAGCCAGCTCATCGCGCTGGC 115  
DB 12473 CTGCGACCATCATGACCGGTGCGGTGTCGCGCGGTGAGTCCAGCGCGGTG 12414  
QY 116 CGCATGCCGAAACACCGGCCAGGATCATGTCAGACAGCGGATGCGGCTTCGATCTC 175  
DB 12413 CAGCGCGCGGTAGTGGAGTTGGTATGTTGCTGTCGGCGAAACCGCCCATCAGGCC 12354  
QY 176 GACGGTGGCGGAGTCCATCTCGTCGCGCGGTGATGTCGCGGAGTGGGCTCAGGACCTC 235  
DB 12353 GCGCGGTTCAGGAGTTGACGCGCGGGTGTGATGCGGAGCGCGGAGTGGCGGT 12294  
QY 236 TGGCAGCGGAGCGCCAGCGGAGACAGTCTGTCAGGCTGGGCTCAGGACCTC 289  
DB 12293 GCGCTCGGAGCAGCTCCAGGATCAGACGTCGCGCGCGGCGGCGGAGCTC 12240

RESULT 20  
US-09-252-991A-6994/c  
; Sequence 6994, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6994  
; LENGTH: 1125  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6994

Query Match  
Best Local Similarity 8.6%; Score 44.8; DB 4; Length 1125;

Best Local Similarity 48.4%; Pred. No. 0.9;  
Matches 155; Conservative 0; Mismatches 162; Indels 3; Gaps 1;  
QY 4 GCAGCAGAACGACCGGTGTCGCGCGCGCGCCATCAGAACTGCCATCTCGCGGGG 63  
DB 767 GCGCCACCATCAGCAAGTCTCCAGGCCGATGCCATGCCGCGCTCAGCGGGC 708  
QY 64 TGACCAGTTCGACCGAGTGGAGTTGACGACGAGCTCATCGCGCGCTCGCGCGGATGC 123  
DB 707 TGCTCAGCGCGACCGGACCAAGTACAGCAGCAAGGCGCAT--CCACAGCGCGAGCCACT 651  
QY 124 CGAAACACCGCGCGAGTACGTCGAGCAGCGGATGCGCGCTCGATCTCAGCGTGC 183  
DB 650 CGAAAGTCGGCGGATGCGCGCAACAGCAGCGCGCGCGAGATCAGGATCGCCA 591  
QY 184 GCGCGAGTTCGATCTCGTCGCGCGGTCTCCACACCGAGGAAATCGCTCGGTGGCAGCG 243  
DB 590 GCAGCAGCGCGCGAGGAGTTCTGCGCGCGCGAGCAGCGGAGGAGTGCATCA 531  
QY 244 GCAGCGCGCGCGGAGTTCGTCAGGCTGCGGTTCAGGCTCAGGACTCTTCGCGGGTTCGGAG 303  
DB 530 GCGAGCATCATCAGCGCGAGCAGGAGTGTGTGAGGGCGCGCTTCAGGCGCTGCTGA 471  
QY 304 CGAAACCGCGCAGCGCTCAG 323  
DB 470 TCCACAGCGCGCGCGAG 451

RESULT 21  
US-09-252-991A-7017/c  
; Sequence 7017, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7017  
; LENGTH: 1929  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7017

Query Match  
Best Local Similarity 48.4%; Pred. No. 0.91;  
Matches 155; Conservative 0; Mismatches 162; Indels 3; Gaps 1;  
QY 4 GCAGCAGAACGACCGGTGTCGCGCGCGCGCCATCAGAACTGCCATCTCGCGGGG 63  
DB 1252 GCGCCACCATCAGCAAGTCTCCAGGCCGATGCCATGCCGCGCTCAGCGGGC 1193  
QY 64 TGACCAGTTCGACCGAGTGGAGTTGACGACGAGCTCATCGCGCGCTCGCGCGGATGC 123  
DB 1192 TGCTCAGCGCGACCGGACCAAGTACAGCAGCAAGGCGCAT--CCACAGCGCGAGCCACT 1136  
QY 124 CGAAACACCGCGCGAGTACGTCGAGCAGCGGATGCGCGCTCGATCTCAGCGTGC 183  
DB 1135 CGAAAGTCGGCGGATGCGCGCAACAGCAGCGCGCGCGAGATCAGGATCGCCA 1076  
QY 184 GCGCGAGTTCGATCTCGTCGCGCGGTCTCCACACCGAGGAAATCGCTCGGTGGCAGCG 243  
DB 1075 GCAGCAGCGCGCGAGGAGTTCTGCGCGCGCGAGCAGCGGAGGAGTGCATCA 1016  
QY 244 GCAGCGCGCGCGGAGTTCGTCAGGCTGCGGTTCAGGACTCTTCGCGGGTTCGGAG 303  
DB 1015 GCGAGCATCATCAGCGCGAGCAGGAGTGTGTGAGGGCGCGCTTCAGGCGCTGCTGA 956

QY 304 CGGAACGCGCAGCGCTCAG 323  
Db 955 TCCACAGCGCGCGCGAG 936

## RESULT 22

US-09-252-991A-7111  
Sequence 7111, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252.991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 7111

LENGTH: 1947

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-7111

Query Match 8.6%; Score 44.8; DB 4; Length 1947;  
Best Local Similarity 48.4%; Pred. No. 0.91;  
Matches 155; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

QY 4 CGACGACGACGACGCTGCTGCGCCCGCGCCATCAAGAACTGCGCACTCCGCGCGGG 63  
Db 636 GCGCCACCATCAGCAAGTCTCCAGCGCATGCCATGCGGATCCGCGGCTCAGCGGGC 755  
QY 64 TGACAGGTTCGACGAGTGGAGTTGAGCAGCAGCTCATCGCGCCTGCGCGGCTATGC 123  
Db 756 TGCTCAGCGGACCGCGACCACTACAGCAGCAAGCGCAT--CCACAGCGCCAGCCACT 812  
QY 124 CGAACACGGGCGAGGATCAGTGCAGCAGCGGATCGCGCTCGATCTCGAGGTG 183  
Db 813 CGAATCGGCGGATCGCGGCAACAGCAGCAGCGCGCGGATCAGGATCGCCCA 872  
QY 184 GCGCAGGTTCGATCTGCTGCGCGCGCTCCACACAGCGGGAATCTGCGTCCGTCGCGAGCG 243  
Db 873 GCAGCAGCGCGCGAGGAGTCTGCGCGCGCAGCAGCGGGAAGTCCCATCA 932  
QY 244 GCAGCCCGAGCGGAGACAGTCTGTCAGGCTGGCGTACGGAACCTCTCGCGGTCGGAG 303  
Db 933 GCGAGCACATCACGCGGACCAAGCAGCAGGCTGCTGAGGCGCGCTTCCAGGCGTGTGA 992  
QY 304 CGGAACGCGCAGCGCTCAG 323  
Db 993 TCCACAGCGCGCGCGAG 1012

## RESULT 23

US-09-616-289-45/c

Sequence 45, Application US/09616289

Patent No. 6632923

GENERAL INFORMATION:

APPLICANT: Lees, Ann M.

APPLICANT: Lees, Robert S.

APPLICANT: Law, Simon W.

APPLICANT: Arjona, Anibal A.

TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING

TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING

FILE REFERENCE: 10797-004001

CURRENT APPLICATION NUMBER: US/09/616,289

CURRENT FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: US 09/517,849

PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 08/979,608  
PRIOR FILING DATE: 1997-11-26  
PRIOR APPLICATION NUMBER: US 60/031,930  
PRIOR FILING DATE: 1996-11-27  
PRIOR APPLICATION NUMBER: US 60/048,547  
PRIOR FILING DATE: 1997-06-03  
NUMBER OF SEQ ID NOS: 53  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 45  
LENGTH: 1614  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1614)  
US-09-616-289-45

Query Match 8.6%; Score 44.6; DB 4; Length 1614;  
Best Local Similarity 47.4%; Pred. No. 0.99;  
Matches 172; Conservative 0; Mismatches 184; Indels 7; Gaps 1;

QY 154 CGCGATGCGCGCTCGATCTCGACGCTCGCGCGCGAGCTCGATCTCGTCCGCCCGGCTCCC 213  
Db 727 CGGCCCGCACCGCGCCCTCCGCGCGGCTGTGCTGCGCGCGCGGCTGCTGCT 668  
QY 214 ACACAGGGGAACTGCTCGTGGTGGCAGCGGAGCCCGCAGCCGCGGACAGCTCGTCCAGGC 273  
Db 667 GCTGTGCGCGCGCGCGCGCTGTGGCGCGCGCGGAGCGCGGCTCCCGCGCGCGACGG 608  
QY 274 TGGCGTCACGAGCTCTCGCGGCTCGGAGCGGAAACGCGCACGGCTCAGATCCCTGTCA 333  
Db 607 CGGGCGGGGGCG 548  
QY 334 GTGCGATCGGCTAGTTCGCGCTGCTCCCTTGGGCTGGGAGATAGCGTTACAGACA 393  
Db 547 CGGGCG 495  
QY 394 GCGGACACG 453  
Db 494 GCGGGCGCAGCGAGGCG 435  
QY 454 TCGGGGCTCGGGCGGTTCGCGCGAGATCCGTACCGCGCGCGCGCGCGCGCGCGCGCG 513  
Db 434 GGGGCG 375  
QY 514 GGG 516  
Db 374 GCG 372

## RESULT 24

US-09-252-991A-1273

Sequence 1273, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 1273

LENGTH: 1839

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-1273

Query Match 8.6%; Score 44.6; DB 4; Length 1839;  
Best Local Similarity 49.0%; Pred. No. 0.99;  
Matches 119; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 15 GACCGTGGTGGCCCGCCGATCAGCAAGTGCACATCGCGGGGTGACCAAGTGG 74  
|||  
Db 1496 GAGAGGAAGTGGTGGCCCGCCGCTCCAGGCTTCACAGAGGCTTCGTTGGTGGG 1555  
|||

QY 75 ACCAGTGGAGTGGAGCAGCAGCTCATCGCGGCTGCGCGGATGCGGAAACACCGG 134  
|||  
Db 1556 GAGGTGAGGAAGATGATCGGCACCAACTCTCTCGCGGCTGTGCTTGATCGCGGGCA 1615  
|||

QY 135 GCCAGATCATGTCAGCAGCGGATGCGGCTCGATCTCGACGCTCGCGCAGCTCG 194  
|||  
Db 1616 GCTCGAGCATCATCAGCGGATCATCATGCTCATGAGACAGTTGCGCGCTCG 1675  
|||

QY 195 ATCTGTCGCCCGGCTCCACACAGCGGAAATGCTCGGTGGCAGCGGACCGCCAGC 254  
|||  
Db 1676 GTCTGAACAGCGCCACCGCTCCAGGCGGTTCGCGGCGTTCAGCAGCGGATGCGCCTGG 1735  
|||

QY 255 CGG 257  
|||  
Db 1736 CGG 1738  
|||

RESULT 25  
US-09-252-991A-1333/c  
; Sequence 1333, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCES: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 1333  
; LENGTH: 1980  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1333

Query Match 8.6%; Score 44.6; DB 4; Length 1980;  
Best Local Similarity 49.0%; Pred. No. 0.99;  
Matches 119; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 15 GACCGTGGTGGCCCGCCGATCAGCAAGTGCACATCGCGGGGTGACCAAGTGG 74  
|||  
Db 599 GAGAGGAAGTGGTGGCCCGCCGCTCCAGGCTTCGACAGCGCTTCGTTGGTGGG 540  
|||

QY 75 ACCAGTGGAGTGGAGCAGCAGCTCATCGCGGCTGCGCGGATGCGGAAACACCGG 134  
|||  
Db 539 GAGGTGAGGAAGATGATCGGCACCAACTCTCTCGCGGCTGTGCTTGATCGCGGGCA 480  
|||

QY 135 GCCAGATCATGTCAGCAGCGGATGCGGCTCATCGAGGTCGCGCGCAGCTCG 194  
|||  
Db 479 GCTCGAAGCATCATCAGCGGATCATCATGCTCATGAGGACCAAGTTGCGCGGCTCG 420  
|||

QY 195 ATCTGTCGCCCGGCTCCACACAGCGGAAATGCTCGGTGGCAGCGGACCGCCAGC 254  
|||  
Db 419 GTCTGAACAGCGCCACCGCTCCAGGCGGTTCGCGGCGTTCAGCAGCGATGCGCCTGG 360  
|||

QY 255 CGG 257  
|||  
Db 359 CGG 357  
|||

RESULT 26

US-09-616-289-48/c  
; Sequence 48, Application US/09616289  
; Patent No. 6632923  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.  
; APPLICANT: Law, Simon W.  
; APPLICANT: Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
; TITLE OF INVENTION: ATHEROSCLEROSIS  
; FILE REFERENCE: 10797-004001  
; CURRENT APPLICATION NUMBER: US/09/616,289  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 09/517,849  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 08/979,608  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 60/031,930  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 60/048,547  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48  
; LENGTH: 2561  
; TYPE: DNA  
; ORGANISM: Oryctolagus cuniculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (246)...(1895)  
US-09-616-289-48

Query Match 8.6%; Score 44.6; DB 4; Length 2561;  
Best Local Similarity 44.7%; Pred. No. 0.99;  
Matches 173; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 130 ACCGGGCGGAGTCACTGTCAGCAGCGGATGCGGCTTCGATTCGACGGTGGCGCA 189  
|||  
Db 507 AGCTGACCGGAGCAGCGCGCTGCTGGATCAGTTTCTCGAGTCCGCGCGCTGGCT 448  
|||

QY 190 GCTGATCTGTCGCCCGGCTCCACACAGGAGGAACTCGCTCGGTGGCAGCGGAGCC 249  
|||  
Db 447 CCGGCTCGGCGCTGCGCGCGCGCACCATTCGGAGATGCGTCCAGTCCGCGCGCG 388  
|||

QY 250 CCAGCGGGAGAGTCTGTCAGGCTGCGCTCAAGGACCTTCGCGGGTCCGAGCGGAAA 309  
|||  
Db 387 CTTTGGCGGAGCGCAGGATGATGTCAGATGTCAGATCCACTCTTGTAGTGGGGAAG 328  
|||

QY 310 CGCGCAGGCTCAGATCCCTGTCAGTCCGCTCAGTCCGCTCAGTCCGCTCCCTTGGCC 369  
|||  
Db 327 CGGCGGAGCAGAGCGCGCGCGCGCGCTGTTGGCGCGCGCGCTCTCGCGCGGG 268  
|||

QY 370 TGGAGGATAGCGGTTTCAAGCAGCGGACCAAGCGCGGGCGGGCGGGGGGTTTCAG 429  
|||  
Db 267 GTAGGCGGGGGCGCGCGCGCTCCGCTCCGCGCGCGCTGACCGTCCGCGCTCCT 208  
|||

QY 430 CCGATCCGCTCAGTACAGCGGCTGCGGGGTGCGGGCGGTGCGGAGATCCGTACCGCC 489  
|||  
Db 207 CCCTCCAGCGCGCGCGCGCGCTCCCGGCTCCCTCCCTCCGCGCTGCGCGCTCCTC 148  
|||

QY 490 CGGACCGGCTCGGCGCAGCGCGCGCGG 516  
|||  
Db 147 CGCCTCCCGCGCGCGCGCGCGCGCG 121  
|||

RESULT 27  
US-09-616-289-50/c  
; Sequence 50, Application US/09616289  
; Patent No. 6632923  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.

```

; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
US-09-410-551B-1

Query Match      8.6%; Score 44.6; DB 4; Length 77536;
Best Local Similarity 44.7%; Pred. No. 1;
Matches 173; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

Qy 10 CGAACACCGGTGGTTCGCCCGCGCGCCATCACGAATGCCACTCCCGCGGGGTGACCA 69
Db 17190 CCATCGCTCCCGCGCGGCAGTCGCGCGGATGACCTGACTCGCAATGCCACACGC 17249

Qy 70 GGTGACACAGTGGAGTTGACAGCAGCTCATCGCCGCTGCGCGGCATCGCGAAC 129
Db 17250 GGGCGCGCTTCCTCGAGCTTGAGGGTTCGCGGCACGACGCGCGCGCATCTCGCCCTGGG 17309

Qy 130 ACCGGGCGCAGATCAGTGTACAGCAGCGGGATGCGCGCCTCGATCTCAGCAGTTCGCGCGCA 189
Db 17310 AGTGTCCGATCACCGGTCGCGCAGACCCCATGCGCTGCCACAGCGCGCCAGGCTCA 17369

Qy 190 GCTCGATCTGTCGCGCGCTCCACACACGAGGGGAACTGGCTGGTGGCAGCGCAGCC 249
Db 17370 CCGCGACCGCCAGCTGGCGCGCTGGACCACTCCACCCGCTCGCGCCACATCCCGGCGCG 17429

Qy 250 CCAGCGCGGACAGCTGTTCCAGGCTGGGCTCACGGACCTCTCGCGGTCGGGAGCGGAAA 309
Db 17430 CCAACATCTCCGCGCATCCAGCGCGGTGTGCGGAGCAACGCCGTGAGCGCACTCTCTCA 17489

Qy 310 CGCGCAGCGCTCAGATCCCTGTGTGTCATCGCATCGGCTCAGTCCGGTGTGTCCTTGCC 369
Db 17490 TAGCGCGCGGAACACCGCGGAGTGGGCACTGAGTTCCACGCCCATCGCGACCTGGG 17549

Qy 370 TGGAGGATAGCGGTTACGACGAGCG 396
Db 17550 CGCCCTGGCGCGGAAGACGACACCG 17576

```

US-08-804-227C-7  
; Sequence 7, Application US/08804227C

; Patent No. 5876991

**GENERAL INFORMATION:**

APPLICANT: DeHoff, Bradley S.

APPLICANT: Kuhstoss, Stuart A.

APPLICANT: Rosteck, Paul R., Jr.

APPLICANT: Sutton, Kimberly L.

; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

; NUMBER OF SEQUENCES: 19

;  
CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501

STREET: LILLY CORPORATE CENTER

; CITY: INDIANAPOLIS

; STATE: IN

; COUNTRY: U

; ZIP: 46285

; COMPUTER READABLE FORM:

```

;
MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM Compatible

```
; OPERATING SYSTEM: MS-DOS
```

```

; SOFTWARE: ASCII (DOS) Text only

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
;
; US-08-804-227C-7
;
; Query Match 8.5%; Score 44.4; DB 2; Length 44377;
; Best Local Similarity 44.8%; Pred. No. 1.1;
; Matches 174; Conservative 0; Mismatches 216; Indels 0; Gaps 0;
;
; QY 8 CACGACGACCGGTGTGTCGCCCGCGCCGATCAGCACTCCCACTCCGGGGGGTGCAC 67
; DB 40450 CCGGCTCCACCCGAGACGGCGCTACCGCCCTGCACAGGCCATCGAGCGGGAAGC 40509
;
; QY 68 CAGGTGACGAGTGGAGTTGAGAGCCAGCTCATCGCGGCTCGCGGGGATGCCGAA 127
; DB 40510 CACGGTCACCGTGGCGGACATCGACTGGGAAACGGTTTCGCCCGCTTCCGCTTCCG 40569
;
; QY 128 ACACCGGGCCAGGATCACGTGCGAGCAGCGGCGCTCGATCTCGACGCTCGGCGG 187
; DB 40570 TCCAGCCCCCTGATCGCGGCACTCCCGGGCCCTACGGCGCCGCGCGCGCGCC 40629
;
; QY 188 CAGCTGATCTGTCGCCCGGCTCCACACCGAGGGAAGAACTGGCTGGTGGAGCGGCG 247
; DB 40630 CGCCGAGGACACCCCGCCCGGCTCTCTGCGGGGCGCGCCCGAGACCGCGCGG 40689
;
; QY 248 CCCAGCGCGGACAGCTCGTCCAGGCTGGCGTCAACGAGCTCTCGCGGGTTCGAGCG 307
;
; Query Match 8.5%; Score 44.4; DB 2; Length 44377;
; Best Local Similarity 44.8%; Pred. No. 1.1;
; Matches 174; Conservative 0; Mismatches 216; Indels 0; Gaps 0;
;
; QY 8 CACGACGACCGGTGTGTCGCCCGCGCCGATCAGCACTCCCACTCCGGGGGGTGCAC 67
; DB 40450 CCGGCTCCACCCGAGACGGCGCTACCGCCCTGCACAGGCCATCGAGCGGGAAGC 40509
;
; QY 68 CAGGTGACGAGTGGAGTTGAGAGCCAGCTCATCGCGGCTCGCGGGGATGCCGAA 127
; DB 40510 CACGGTCACCGTGGCGGACATCGACTGGGAAACGGTTTCGCCCGCTTCCGCTTCCG 40569
;
; QY 128 ACACCGGGCCAGGATCACGTGCGAGCAGCGGCGCTCGATCTCGACGCTCGGCGG 187
; DB 40570 TCCAGCCCCCTGATCGCGGCACTCCCGGGCCCTACGGCGCCGCGCGCGCGCC 40629
;
; QY 188 CAGCTGATCTGTCGCCCGGCTCCACACCGAGGGAAGAACTGGCTGGTGGAGCGGCG 247
; DB 40630 CGCCGAGGACACCCCGCCCGGCTCTCTGCGGGGCGCGCCCGAGACCGCGCGG 40689
;
; QY 248 CCCAGCGCGGACAGCTCGTCCAGGCTGGCGTCAACGAGCTCTCGCGGGTTCGAGCG 307
; DB 40690 GTTCGCCCTTGGACCTGCTGCTCCGCGACATCGTCGCGGCGGTCTTGGCCCACTCCGAGACGC 40749
;
; QY 308 AACCGGACAGGCTCAGATCCCTGTGATGCGCATCGGCTCAGTCCGCGGTCTGCCCTTGG 367
; DB 40750 CCGGGTCGACGCGCCGCGCCCTTCCGGAACCTCGGCTTCGACTCGCTCGCGCGGTGCG 40809
;
; QY 368 CTGGGAGATAGCGTTACACAGCGG 397
; DB 40810 GTTCGCCCGCGGCTGGCGCGAGACACCGG 40839
;
; RESULT 30
; US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.

```



Db 40690 GCTGCCCTGAGCTGCTCCGCGCAGCTGCGCGGCTCTCGGCGCACTCCGAGGACGC 40749  
QY 308 AACGGCAGCGCTCAGATCCCTGTCAGTCGATCGGCTCAGTCGCGGTCTGTCCTCCCTGG 367  
Db 40750 CCGGTCAGACCGCGGCGGCTCCGCGGACTCGGCTTCGATCGCTCGCGCGGTGCG 40809  
QY 368 CTGGAGGATAGCGGTTACGACGAGCGG 397  
Db 40810 GCTGCGCGCGGCTGGCGGAGACACCGG 40839

## RESULT 31

US-09-252-991A-13718/c  
; Sequence 1, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13718  
; LENGTH: 927  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13718

Query Match 8.5%; Score 44.2; DB 4; Length 927;  
Best Local Similarity 48.6%; Pred. No. 1.2;  
Matches 121; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 32 GCCGCCCATCACGAATGCCACTCCGCGGGGTGACCAAGTCCACAGGTGGGAGTTGAG 91  
Db 384 GCGCGCATCACTCCGCGATGCTGCTCGGTGAATACATCATCGATCGGCGCGGTGTC 325  
QY 92 CAGCAGTCTATCCGCGCTGCGCGCGCATGCCGAAACACCGCGCGCAGGATCACTGTCAG 151  
Db 324 GCGGCTACACGAGGAGACCTGCTCCATCAGCTCGCGCGCGGAGACTCCGCTGTCATCAG 265  
QY 152 CACGCGATCGCGCTCGATCTCGACGTCGCGCGCGAGCTCGATCTCGTCGCGCGCTC 211  
Db 264 GATCAGTCGCGCGAGCGGCTTCGACGCGCGCGCGGAGTCTGATGTCAGCCCTGACTC 205  
QY 212 CCACACGAGGAACTGGTGGTGGCGAGCGCGCGAGCCCGAGCGGACAGCTCGTCCAG 271  
Db 204 GTCCACCACTCGAAACCGCTCTCGAGCAGGCGGCGGCGCTCGAGGCTCGGACTTCTTCGG 145  
QY 272 GCTGCGCTC 280  
Db 144 GGTGCTC 136

## RESULT 32

US-09-105-537-1/c  
; Sequence 1, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-1

Query Match 8.5%; Score 44.2; DB 3; Length 15872;  
Best Local Similarity 45.8%; Pred. No. 1.2;  
Matches 189; Conservative 0; Mismatches 223; Indels 1; Gaps 1;

QY 32 GCGGCCCATCACGAATGCCACTCCGCGGGGTGACCAAGTCCACAGGTGGGAGTTGAG 91  
Db 2191 GCGGTGCGAGACGCTGAGCGGCTTCGTCGCGCGGCTGTTCCGCGAGCAGCTCGCGAC 2132  
QY 92 CAGCCAGCTCATCGCGCTGCGCGGATCGCGAAACACACCGGCGCAGATCACGTGAG 151  
Db 2131 GTCGAGGAGCGCGCTGCGCGCGCGCGCGAGGACGACCGCGCGGGCGCTTCCGCGCGC 2072  
QY 152 CAGCGGATGCGCGCTCGATCTCGACGTCGCGCGCGAGCTCGATCTCGTCGCGCGGCTC 211  
Db 2071 GAGGAGAGCTCCGCTCGCGTCCGCGAGGTGCGAAGCACTTCTCTGCTCGCTC 2012  
QY 212 CCACACCGGGAATGCTCGGTGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 271  
Db 2011 GACCGGACCATCGCGCGCGCTCGGAGAGCGCTGCAATGAGCGCGCGCGCGCGCGCG 1952  
QY 272 GCTGCGTACGAGACTCTCGCGGTGCGGAGCGGAAACCGCGCACGGCTCAGATCCCTGT 331  
Db 1951 GAGCGCGCGCGCTCGCGCAGCAGGACCGCGCGCGCGCGCGCTGATCTC- 1893  
QY 332 CAGTCGATCGCTCAGTCGCGGTGCTCCCTTGGCTGGGAGGATAGCGGTTCCAGAC 391  
Db 1892 CGACGAGTGGCGCGCGAGCTCGGCGTGAOCCCGCGAGACTCGACGAGCGGTGGA 1833  
QY 392 GAGCGCGACCGCGCGCGCGCGCGGCGGCTTCAGCGATCGGCTCGATG 444  
Db 1832 GGGCGACCTCCACGCGGAGAGCGGCGCTGTGTGACGGTGGGTCGAGG 1780

## RESULT 33

US-09-091-609-1/c  
; Sequence 1, Application US/09091609  
; Patent No. 6600029  
; GENERAL INFORMATION:  
; APPLICANT: SHERMAN, DAVID H.  
; APPLICANT: WILLIAMS, MARK D.  
; APPLICANT: XUE, YONGQUAN  
; TITLE OF INVENTION: METABOLIC ENGINEERING OF  
; FILE REFERENCE: 600.297US2  
; CURRENT APPLICATION NUMBER: US/09/091,609  
; CURRENT FILING DATE: 1998-06-19  
; EARLIER APPLICATION NUMBER: PCT/US96/20119  
; EARLIER FILING DATE: 1996-12-18  
; EARLIER APPLICATION NUMBER: 60/008,847  
; EARLIER FILING DATE: 1995-12-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (20)....(13909)  
US-09-091-609-1

Query Match 8.5%; Score 44.2; DB 4; Length 15872;  
Best Local Similarity 45.8%; Pred. No. 1.2;  
Matches 189; Conservative 0; Mismatches 223; Indels 1; Gaps 1;

QY 32 GCGGCCCATCACGAATGCCACTCCGCGGGGTGACCAAGTCCGAGGTCGAGGTCGAGGTTGAG 91

Db 2191 GCGTGCAGACGCTGAGCCGCTTCGTCGCGCGCCCTGTCGCGAGAGCTCGCGAC 2132  
QY 92 CAGCCAGCTCATCGCCCTGCGCGCATGCCGAAACCGGCGCAGATCACGTCAG 151  
Db 2131 GTCGAGAGACGCGCGCTCGCGCCCGGAGACGACCGCGCGCGCCGTTACGCGCCG 2072  
QY 152 CAGCGGCATCGCGCCCTCGATCTCGAGCGTCCGCGCAGCTCATCTCGTCGCGCGGCTC 211  
Db 2071 GAGGAGAGCTCCGCTCGGTCGCGCGAGGTCGGAAGCACTTCCTCTCGTCGCTC 2012  
QY 212 CACACAGGAGAACTGGGTCGGTGCAGCGCAGCGCAGCCCGAGCGGAGATCGTCAG 271  
Db 2011 GACCGGACCATCGCGCGCCCTCGGAGAGCGCTCATGAGCGCGCCGCGCGCGAC 1952  
QY 272 GCTGGGCTCACGACCTCTCGCGGTCGAGCGGAAACCGCACCGCTCAGATCCCTGT 331  
Db 1951 GAGCGGGCGCGCTCGCGAGACACAGGACCCCGGCGAGTGCGCGCGCTGATCTC-GC 1893  
QY 332 CAGTCGATCGGTCAGTCGCGTTCGCTGCGGTCGCGGTCGCGGAGATCGGTTACGAC 391  
Db 1892 CGACGGAGTGGCGGCGAGCAGGTCGCGGCTGACGCGCCCGAGGACTCGACGAGCGGTGA 1833  
QY 392 GAGCGGACCAACGCGCGCGCGGCGGCGGTTACGCGGATCCGCTCGATG 444  
Db 1832 GGGCGACCTCACGGCGAGAGCGCGGCTGTGTGTGAGCGGTGCGGTGAGG 1780

## RESULT 34

US-09-091-609-3/c  
; Sequence 3, Application US/09091609  
; Patent No. 6600029  
; GENERAL INFORMATION:  
; APPLICANT: SHERMAN, DAVID H.  
; APPLICANT: WILLIAMS, MARK D.  
; APPLICANT: XUE, YONGQUAN  
; TITLE OF INVENTION: METABOLIC ENGINEERING OF  
; FILE REFERENCE: 600.297US2  
; CURRENT APPLICATION NUMBER: US/09/091,609  
; CURRENT FILING DATE: 1998-06-19  
; EARLIER APPLICATION NUMBER: PCT/US96/20119  
; EARLIER FILING DATE: 1996-12-18  
; EARLIER APPLICATION NUMBER: 60/008,847  
; EARLIER FILING DATE: 1995-12-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (14148)...(15824)  
US-09-091-609-3

Query Match 8.5%; Score 44.2; DB 4; Length 15872;  
Best Local Similarity 45.8%; Pred. No. 1.2;  
Matches 189; Conservative 0; Mismatches 223; Indels 1; Gaps 1;

QY 32 GCGCGCCATCAGCACTGCCACTCCGCGGGGTGACCAAGTGCACAGTGGAGTTGAG 91  
Db 2191 GCGTTCGAGACGCTGAGCCGCTTCGTCGCGCGCCCTTCGCGCAGAGTCGCGCAG 2132  
QY 92 CAGCCAGCTCATCGCGCCCTGCGCGCATGCCGAAACCGGCGCAGATCACGTCGAG 151  
Db 2131 GTCGAGAGCGCGCTCGCGCGCCCGGAGGACGACCGCGCGCGCGCTTCACGCGCG 2072  
QY 152 CAGCGGATCGCGGCTCGATCTCGAGCGTCGCGCGCAGCTGATCTCGTCGCGGCTC 211  
Db 2071 GAGGAGAGCTCCCGCTCGGTCGCGGAGGTCGGAAGCACTTCCTCTCGCTCGCCTC 2012  
QY 212 CACACAGGAGAACTGGGTCGTCGTCGAGCGGAGCGCGCCAGCGCGGACAGCTCGTCCAG 271

Db 2011 GACCGGACCATCGCGCGCCCTCGGAGAGCGCTCATGAGCGCGCGCGCGAC 1952  
QY 272 GCTGCGCTCACGACCTCTCGCGGGTCGGAGACGCGAAACCGCACGCGCTCAGATCCCTGT 331  
Db 1951 GAGCGCGCGCGCTCGCGAGACGACCGCGCGCGCTGCGGCTGATCTC-GC 1893  
QY 332 CAGTCGATCGGTCAGTCGCGGTCGCTGCGGTCGCGGAGATCGGTTACGAC 391  
Db 1892 CGACGGAGTGGCGCGCGGAGAGGTCGCGGCTGACGCGCCCGAGGACTCGACGAGCGGTGA 1833  
QY 392 GAGCGGACCAACGCGCGCGCGGCGGCGGTTACGCGGATCCGCTCGATG 444  
Db 1832 GGGCGACCTCACGCGAGAGCGCGGCTGTGTGTGAGCGGTGCGGTGAGG 1780

## RESULT 35

US-08-658-136-2/c  
; Sequence 2, Application US/08658136  
; Patent No. 6071717  
; GENERAL INFORMATION:  
; APPLICANT: KLINGER, KATHERINE W  
; APPLICANT: LANDES, GREGORY M  
; APPLICANT: BURN, TIMOTHY C  
; APPLICANT: CONNORS, TIMOTHY D  
; APPLICANT: DACKOWSKI, WILLIAM  
; APPLICANT: GERMINO, GREGORY  
; APPLICANT: QIAN, FENG  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENZYME CORPORATION  
; STREET: ONE MOUNTAIN ROAD  
; CITY: FRAMINGHAM  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/658,136  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LASSEN, ELIZABETH  
; REGISTRATION NUMBER: 31,845  
; REFERENCE/DOCKET NUMBER: GEN4-17.8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 508-872-8400  
; TELEFAX: 508-872-5415  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 53526 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-658-136-2

Query Match 8.5%; Score 44; DB 3; Length 53526;  
Best Local Similarity 44.4%; Pred. No. 1.3;  
Matches 176; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

QY 94 GCGAGCTCATCGCGCCCTCGCGCGCATCGCGAAACCGGCGCAGATCACGTCAGCA 153  
Db 3682 GCGAGCCAGCGCGCGCGCGCGCATCGTTAGGCGCGCGCGCATGCCCC 3623  
QY 154 CGGCGATGCGCGCTTCGATCTCGACGCTCGCGCGCGCAGCTCGATCTCGTCGCGGCTCC 213  
Db 3622 CGCGTCCCGAGCGCCCGCGCGGAGCGCGAGCTCAGGCGGCGCGCGGAGCGC 3563



```
RESULT 38
US-09-105-537-3/c
; Sequence 3, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-3
Query Match      8.4%; Score 43.6; DB 3; Length 13613;
Best Local Similarity 51.0%; Pred. No. 1.6;
Matches 103; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 314 CACGGCTCAGATCCCTGTGTCAGTCGCATCGGCTCAGTCGCGGTCGTCGCCCTTGGCCCTGGG 373
DB 3473 CAGGGGTGAGGCTCGGGGCGAGGGGTGCGGCTCACTGCGGGGTGCGCGCGGCGCGGC 3414
QY 374 AGGATAGCGGTTACAGACGAGCGGACACGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 433
DB 3413 GGTGTGCGCGCGAGCGGCTCCAGCTCGGGGACGATCCCGCGGGGGTGGGGTGGCGGA 3354
QY 434 TCCGCTCGATGACACGCGGCTCGGGGTCGGGGCGGTGCGGACGATCCGTCACGCCCGGA 493
DB 3353 AGGTCTCTCGCGGACGCGGTGCGGGCGGTGGCGACGAGGGGTGCTCGAGGATGCGGA 3294
QY 494 CGGCTCGCGGACGCGCGCGG 515
DB 3293 CGACGGGCTCCGCGACGGGCTG 3272

RESULT 39
US-09-320-878-19/c
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
```

```
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19
Query Match      8.4%; Score 43.6; DB 3; Length 38506;
Best Local Similarity 51.0%; Pred. No. 1.6;
Matches 103; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 314 CACGGCTCAGATCCCTGTGTCAGTCGCATCGGCTCAGTCGCGGTCGTCGCCCTTGGCCCTGGG 373
DB 37470 CAGGGGTGAGGCTCGGGGCGAGGGGTGCGGCTCACTGCGGGGTGCGCGCGGCGCGGC 37411
QY 374 AGGATAGCGGTTACAGACGAGCGGACACGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 433
DB 37410 GGTGTGCGCGCGAGCGGCTCCAGCTCGGGGACGATCCCGCGGGGGTGGGGTGGCGGA 37351
QY 434 TCCGCTCGATGACACGCGGCTCGGGGTCGGGGCGGTGCGGACGATCCGTCACGCCCGGA 493
DB 37350 AGGTCTCTCGCGGACGCGGTGCGGGCGGTGGCGACGAGGGGTGCTCGAGGATGCGGA 37291
QY 494 CGGCTCGCGGACGCGCGCGG 515
DB 37290 CGACGGGCTCCGCGACGGGCTG 37269

RESULT 40
US-09-141-908-1/c
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-1
Query Match      8.4%; Score 43.6; DB 4; Length 38506;
Best Local Similarity 51.0%; Pred. No. 1.6;
Matches 103; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 314 CACGGCTCAGATCCCTGTGTCAGTCGCATCGGCTCAGTCGCGGTCGTCGCCCTTGGCCCTGGG 373
DB 37470 CAGGGGTGAGGCTCGGGGCGAGGGGTGCGGCTCACTGCGGGGTGCGCGCGGCGCGGC 37411
QY 374 AGGATAGCGGTTACAGACGAGCGGACACGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 433
DB 37410 GGTGTGCGCGCGAGCGGCTCCAGCTCGGGGACGATCCCGCGGGGGTGGGGTGGCGGA 37351
QY 434 TCCGCTCGATGACACGCGGCTCGGGGTCGGGGCGGTGCGGACGATCCGTCACGCCCGGA 493
DB 37350 AGGTCTCTCGCGGACGCGGTGCGGGCGGTGGCGACGAGGGGTGCTCGAGGATGCGGA 37291
```

QY 494 CCGCTCGGCGGCGGCGGCGG 515  
Db 37290 CGAGGCGTCCCGCAGGCGTG 37269

## RESULT 41

US-09-657-440-19/c  
; Sequence 19, Application US/09657440  
; Patent No. 6509455  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120  
; CURRENT APPLICATION NUMBER: US/09/657,440  
; CURRENT FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patent In. Ver. 2.0  
; SEQ ID NO 19:  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-657-440-19

Query Match 8.4%; Score 43.6; DB 4; Length 38506;  
Best Local Similarity 51.0%; Pred. No. 1.6; Indels 0; Gaps 0;  
Matches 103; Conservative 0; Mismatches 99;  
QY 314 CAGCGTCAAGTCCCTGTCAGTGCATCGGCTCAGTCCGCGTCTGCTCCCTTGGCGTGGG 373  
Db 37470 CAGGGGTGAGGCTGGGGCGAGGGGTGCGGCTCAGTCCGCGGCTGCGCGCGGCGGCGG 37411  
QY 374 AGATAGCGGTTCACACAGGCGGCACACGCGGGCGGGCGGGCGGGCGGGTTCAGCCGA 433  
Db 37410 GGTGTGCGCGGCGAGCGGCTCCAGCTCGGGGACGATCCCGCGGGGTGGGGTCCCGGA 37351  
QY 434 TCGCTCGATGACACAGCGGCTCGGGGTGCGGGCGGCTCGGCGAGATCCGTACCGCCCGGA 493  
Db 37350 AGTCTCTCCGCGCAGCGCGTGGCGGGGTGGCGAGGCTGCTCGAGGATCGCGA 37291  
QY 494 CCGCTCGGCGGCGGCGGCGG 515  
Db 37290 CGACGCGTCCCGCAGGCGTG 37269

## RESULT 42

US-09-252-991A-4109  
; Sequence 4109, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4109:  
; LENGTH: 366  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4109

Query Match 8.3%; Score 43.4; DB 4; Length 366;  
Best Local Similarity 47.6%; Pred. No. 1.7;  
Matches 128; Conservative 0; Mismatches 141; Indels 0; Gaps 0;  
QY 11 GAAGGACCGGTGCTCGCCCGCGCCCATCACAACTGCCACTCCCGCGGGGTGACCAG 70  
Db 63 GATCGCCACCTGCTCGCGCTGGAGCGGAGCCACCTCTCGGACATTTCTCTGGCCATCAT 122  
QY 71 GTCACACAGGTGGAGTTGAGCAGCAGCTCATGCGCCCTGCGCCGCGCATGCCGAAACA 130  
Db 123 GGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182  
QY 131 CCGGCGCAGGATCACGTGACGAGCGGATGCGGCGCTCGATCTCGACGCTCGGCGCAG 190  
Db 183 CAGCTCACTTCTGCTCTCGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 242  
QY 191 CTCGATCTGCTCGCCCGCTCCACACAGGGGAACTGGCTCGGTGGCAGCGGCGAGCCC 250  
Db 243 CTTGGCGATGCTCTGCTCACCCCTGGCTGGCAGGGGCTCGGCTACTTCAACAACTA 302  
QY 251 CAGCGGCGACAGCTCGTCCAGGCTGGCGT 279  
Db 303 CACCTGGCGCGCTGCTCTCTCTAGTGT 331

## RESULT 43

US-09-489-039A-123  
; Sequence 123, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 123:  
; LENGTH: 1287  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-123

Query Match 8.3%; Score 43.4; DB 4; Length 1287;  
Best Local Similarity 48.6%; Pred. No. 1.7;  
Matches 119; Conservative 0; Mismatches 126; Indels 0; Gaps 0;  
QY 40 TCAGGACTGCCACTCGCGGGGTGACAGGTGACAGGTGACAGGTGAGGTGAGGAGCAGC 99  
Db 672 TGAGGATCGCCCGCAGCAGGTGGGCTGCTGCCAGATATTGTGGAACGTGCCAGCGACA 731  
QY 100 TCATCGCGGCTGCGCGCGCATGCCGAAACACCGGCGCCAGGATCAGTGCAGCACCGCGA 159  
Db 732 GCTCATCCACAGCGACACGCCGACGCCAGTTGGAGATAAAGCGCGCAGCGCCAGCG 791  
QY 160 TGGCGGCTCGATCTCGACGCTCGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 219  
Db 792 CGAGCAGCGGTACGCGGATGGTCAAGCTCAGCCAGGCGCCACCAACCGCGCGCCACCC 851  
QY 220 GGGGAACTGGCTCGGTGGCAGCGGCGAGCCCGCCAGCGGACAGCTCGTCCAGGCTGCGGT 279  
Db 852 AGCGCATTTGCCACGCGGAGCGCGCGCGCTTGGGACGCTTGGCCCTTATGCTGATGACAA 911  
QY 280 CACGG 284  
Db 912 AGCGG 916

## RESULT 44

US-08-740-223A-17/c

; Sequence 17, Application US/08740223A  
; Patent No. 6265564  
; GENERAL INFORMATION:  
; APPLICANT: Davis, et al.  
; TITLE OF INVENTION: Expressed Ligand - Vascular  
; TITLE OF INVENTION: Intercellular Signalling Molecule  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
; STREET: 777 Old Saw Mill Road  
; CITY: Tarrytown  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/740,223A  
; FILING DATE: 25-OCT-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: USSN 60/022/999  
; FILING DATE: 02-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cobert, Robert J  
; REGISTRATION NUMBER: 36,108  
; REFERENCE/DOCKET NUMBER: REG 333  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 914-345-7400  
; TELEFAX: 914-345-7721  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1512 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 1...1509  
; OTHER INFORMATION:  
; NAME/KEY: TIE ligand-4  
; LOCATION: 1...1512  
; OTHER INFORMATION:  
; US-08-740-223A-17

Query Match 8.3%; Score 43.4; DB 3; Length 1512;  
Best Local Similarity 49.8%; Pred. No. 1.7;  
Matches 110; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
QY 44 GAATGCCACTCCGGCGGGTGCACAGTCCAGCGTGGAGTTGAGCCAGCTCAT 103  
Db 564 GAGCTTCTGCTCTGTAGCAGAGTGTCTCCAGCTTGTGTGGACAGAAAGGTCTC 505  
QY 104 CGCGCCCTCGCGCATGCCGAAACACCGGCGCAGGATCAGTGCAGCAGCGGATGCG 163  
Db 504 TGGCATCTGGGATCATCTTGTATGTTCTGGTTAGGAGCTGAGCCTCATGTCGTCAG 445  
QY 164 CGCTCGATCTCGACGGTGGCGGCGAGCTCGATCTGTCGCCGGCTCCACACAGGGG 223  
Db 444 CTTGGGATCTGGCAGTGTCTGTTTCAAGAGGCTGTGCGCCAGCTCTAGCATGGGGC 385  
QY 224 AAATGGCTCGGTGGCAGCGGCGAGCCCGGAGGAGCT 264  
Db 384 CGTCTGATCTGGCCATTTGCTGTGACCTGTCTCCAGCT 344

RESULT 45  
US-09-202-491-9/c  
; Sequence 9, Application US/09202491

; Patent No. 6432667  
; GENERAL INFORMATION:  
; APPLICANT: Valenzuela et al.  
; TITLE OF INVENTION: NOVEL LIGANDS, METHODS OF MAKING AND USES THEREOF  
; FILE REFERENCE: REG330-K  
; CURRENT APPLICATION NUMBER: US/09/202,491  
; CURRENT FILING DATE: 1998-11-16  
; EARLIER APPLICATION NUMBER: PCT/US97/10728  
; EARLIER FILING DATE: 1997-06-19  
; EARLIER APPLICATION NUMBER: 60/022,999  
; EARLIER FILING DATE: 1996-08-02  
; EARLIER APPLICATION NUMBER: 60/021,087  
; EARLIER FILING DATE: 1996-07-02  
; EARLIER APPLICATION NUMBER: 08/665,926  
; EARLIER FILING DATE: 1996-06-19  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 1512  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1509)  
; US-09-202-491-9

Query Match 8.3%; Score 43.4; DB 4; Length 1512;  
Best Local Similarity 49.8%; Pred. No. 1.7;  
Matches 110; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
QY 44 GAATGCCACTCCGGCGGGTGCACAGTTCGACAGGTGGAGTTGAGCCAGCTCAT 103  
Db 564 GAGCTTCTGCTCTGTAGCAGAGTGTCTCCAGCTTGTGTGGACAGAAAGGTCTC 505  
QY 104 CGCGCCCTCGCGCATGCCGAAACACCGGCGCAGGATCAGTGCAGCAGCGGATGCG 163  
Db 504 TGGCATCTGGGATCATCTTGTATGTTCTGGTTAGGAGCTGAGCCTCATGTCGTCAG 445  
QY 164 CGCTCGATCTCGACGGTGGCGGCGAGCTCGATCTGTCGCCGGCTCCACACAGGGG 223  
Db 444 CTTGGGATCTGGCAGTGTCTGTTTCAAGAGGCTGTGCGCCAGCTCTAGCATGGGGC 385  
QY 224 AAATGGCTCGGTGGCAGCGGCGAGCCCGGAGGAGCT 264  
Db 384 CGTCTGATCTGGCCATTTGCTGTGACCTGTCTCCAGCT 344

RESULT 46  
US-09-709-188-17/c  
; Sequence 17, Application US/09709188  
; Patent No. 6441137  
; GENERAL INFORMATION:  
; APPLICANT: Davis et al.  
; TITLE OF INVENTION: Expressed Ligand - Vascular Intercellular Signaling Molecule  
; FILE REFERENCE: REG 333-Z  
; CURRENT APPLICATION NUMBER: US/09/709,188  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 08/740,223  
; PRIOR FILING DATE: 1996-10-25  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 17  
; LENGTH: 1512  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1512)  
; OTHER INFORMATION:  
; US-09-709-188-17

Query Match 8.3%; Score 43.4; DB 4; Length 1512;  
Best Local Similarity 49.8%; Pred. No. 1.7;

Matches 110; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 44 GAACCTCCACTCCGCGGGGTGACAGGTCGACAGGTGGAGTTGAGCAGCCAGCTCAT 103  
Db 564 GAGCTTCTGCTCTGAGCAGAGCTGGTCTCCAGCTTGTGGTGACAGAAAGGTTCTC 505

QY 104 CGCGGCTCGCGCGGATGCGGAAACACCGGGCCAGGATCAGTGCAGCAGCGCGATCG 163  
Db 504 TGCCATCTGGGCATCATTCTTGATGCTGTTTTCAGGAGCTGAGCTCCATGCTCGTCA 445

QY 164 CGGCTGATCTGACAGGTGCGCGGAGCTGATGCTGCTGCGCGGTCGCCACAGCGGG 223  
Db 444 CTTGCGGATCTGGCGAGTGTCTGTTTCAGGAGGCTGTGCGCCAGCTCTAGCATGGGGC 385

QY 224 AAACCTGGCTCGGTGGCGAGCGGCGAGCCCGGAGCGGACAGCT 264  
Db 384 CGTCTGATCTGGCCATTGCTGTGACCTGCTCCAGCT 344

RESULT 47  
US-09-489-039A-241/c  
; Sequence 241, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breston et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 241  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-241

Query Match 8.3%; Score 43.4; DB 4; Length 1773;  
Best Local Similarity 48.6%; Pred. No. 1.7;  
Matches 119; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 40 TCACGAACCTGCCACTCCGCGGGGTGACAGGTCGACAGGTGGGAGTTGACGACCCAGC 99  
Db 1090 TGAGGATGCCCGCAGCAGGTTGGGCTGTGCGAGATATTGGAACTGCGCAGCGACA 1031

QY 100 TCTATCCGCTCGCGCGGATGCGGAAACACCGGGCCAGGATCAGTGCAGCAGCGGGA 159  
Db 1030 GCTCATCCACAGCGCACGCGGAGCGCCCGAGTTGGAGATAAAGCGGCGAGCGCCAGC 971

QY 160 TGGCGCTCGATCTCGAGGTCGCGCGGAGCTCGATCTCGTCCGCGCTCCACACCA 219  
Db 970 CGAGCAGCGGTAGCGCGATGTTACGCTCAGCCAGGCCACACCGCCCGCGCACCC 911

QY 220 GGGGAACTGGTTCGGTGCAGCGGAGCGCCCGAGCGGAGCAGCTCGTCCAGGCTGGCGT 279  
Db 910 AGCGCCATTGCGCCAGCGGAGCGCGCGCGCTGGAGAGCCCTTGCCTTCATGGTGACA 851

QY 280 CACGG 284  
Db 850 AGCGG 846

RESULT 48  
US-08-460-269C-3  
; Sequence 3, Application US/08460269C  
; Patent No. 6197549  
; GENERAL INFORMATION:  
; APPLICANT: CLARE, JEFFREY J.  
; APPLICANT: ROMANOS, MICHAEL A.  
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST

NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Millen, White, Zelano & Branigan, P.C.  
STREET: 2000 Clarendon Blvd., Suite 1400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,269C  
FILING DATE: 02-Jun-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lebovitz, Richard M.  
REGISTRATION NUMBER: 37,067  
REFERENCE/DOCKET NUMBER: Popov-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 243-6333  
TELEFAX: (703) 243-6410  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3000 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 145..2877  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-08-460-269C-3

Query Match 8.3%; Score 43.4; DB 3; Length 3000;  
Best Local Similarity 45.4%; Pred. No. 1.7;  
Matches 193; Conservative 0; Mismatches 231; Indels 1; Gaps 1;

QY 71 GTCACACAGGTGGGAGTTGAGCAGCCAGCTCATCGCGCTGCGCGGATGCGGCAATGCCAAACA 130  
Db 2182 GCCGACACACGGTGGCGGTGGCGGCGGCTGGCACTGGCGGCTGCGCGGTAT 2241

QY 131 CGCGGCGGATCAGTGCAGCAGCGGATGCGGCTCGATCTCGACGTCGCGCGCAG 190  
Db 2242 ACGGCGGCGACCGCGGCTTTTACCGGCGACGGCGGCGGCACACCGACAG-CGTGCTGT 2300

QY 191 CTCGATCTCGTCCGCGCTCCACACAGGGGAACTGGCTCGGTGGCAGCGCGAGCCC 250  
Db 2301 CGGGGGCTATGCCACCTATATGCCAACAGCGTTTCTACTGAGCGGACGCTGCGCGC 2360

QY 251 CAGCGGCGGACAGTCTGTCAGGCTGGCGTACCGAATCTTCGCGGGTCGGGAGCGGAAC 310  
Db 2361 CAGCGGCTCGAAATGACTTCAAGGTGGCGGCGAGTGGGTAGCGGTCAAGGGCAA 2420

QY 311 GCGCAGCGCTCAGATCCCTGTCAGTGCATCGGCTCAGTCCGCTCGTCCCGCTTGGCCT 370  
Db 2421 GTACCGCACCCATGGGGTAGGCGCTTCGTCGAGGGGGCGCGCTTCGCCCATGCCGA 2480

QY 371 GGAGAGATACGGTTTACGACAGCGGCGACACCGCGGGCGGGCGGGCGGGTTCAGC 430  
Db 2481 CGGCTGGTTCTCGAGCGCGGAGAGTGGCGGTTCGCGGCTCGGCGCGGTTCGTA 2540

QY 431 CGATCGCTCGATGACGAGCGGCTGGGGTTCGGGCGGAGATCCGTACGCGCC 490  
Db 2541 CCGGCGGCGCAATGGCTCGCGGTTCGCGAGGAGCGGAGCTCGGTGCTGGGTGCCT 2600

QY 491 GGACC 495  
Db 2601 GGGCC 2605

## RESULT 49

US-09-344-510B-6/c  
; Sequence 6, Application US/09344510B  
; Patent No. 6579850  
; GENERAL INFORMATION:  
; APPLICANT: Nabeshima, Youichi  
; Kuroo, Makoto  
; Sekine, Susumu  
; Iida, Akihiro  
; TITLE OF INVENTION: No. 6579850el Polypeptide, No. 6579850el DNA and No. 6579850el  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10112-3801  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 1.44 mb, DS, DD  
; COMPUTER: Compaq DeskPro EN  
; OPERATING SYSTEM: Windows 98  
; SOFTWARE: WordPad  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/344,510B  
; FILING DATE: 25-Jun-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/04585  
; FILING DATE: 12-DEC-1997  
; APPLICATION NUMBER: JP 347871  
; FILING DATE: 26-DEC-1996  
; APPLICATION NUMBER: JP 205815  
; FILING DATE: 31-JUL-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perry, Lawrence S.  
; REGISTRATION NUMBER: 31865  
; REFERENCE/DOCKET NUMBER: 766.32  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 218-2100  
; TELEFAX: (212) 218-2200  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3163  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; IMMEDIATE SOURCE:  
; LIBRARY: kidney  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 9..3047  
; IDENTIFICATION METHOD: E  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-344-510B-6  
Query Match 8.3%; Score 43.4; DB 4; Length 3163;  
Best Local Similarity 53.9%; Pred. No. 1.7;  
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
QY 189 AGCTCGATCTCGTCCCGGCTCCACACACAGGGGAACTGCTCGGTGGCGAGCGGCAGC 248  
DB 574 ACCACGGGCTGACCGCCAGCTCCCGCAGCGCTCCACAGCGCGCGGTAGTAGCGCAGC 515  
QY 249 CCCAGCCGGGACAGCTCGTCCAGGCTGGCGTCCAGGACCTCTCGCGGTCCGGAGCGGAA 308  
DB 514 CCTCGCGGTTGGGACGCGCGCTGCCATTGGGGAGCACTCGCGCCACAGATGGAG 455  
QY 309 ACCGACAGGCTCAGATCCCTGTGCTAGTCGATCGATCGCTCAGTGGCG 353  
DB 454 AAGCGGTAGTAGTACCCCGAGCTCCGCGACGCGCTCCGCTGCG 410

## RESULT 50

US-09-344-510B-7/c  
; Sequence 7, Application US/09344510B  
; Patent No. 6579850  
; GENERAL INFORMATION:  
; APPLICANT: Nabeshima, Youichi  
; Kuroo, Makoto  
; Sekine, Susumu  
; Iida, Akihiro  
; TITLE OF INVENTION: No. 6579850el Polypeptide, No. 6579850el DNA and No. 6579850el  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10112-3801  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 1.44 mb, DS, DD  
; COMPUTER: Compaq DeskPro EN  
; OPERATING SYSTEM: Windows 98  
; SOFTWARE: WordPad  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/344,510B  
; FILING DATE: 25-Jun-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/04585  
; FILING DATE: 12-DEC-1997  
; APPLICATION NUMBER: JP 347871  
; FILING DATE: 26-DEC-1996  
; APPLICATION NUMBER: JP 205815  
; FILING DATE: 31-JUL-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perry, Lawrence S.  
; REGISTRATION NUMBER: 31865  
; REFERENCE/DOCKET NUMBER: 766.32  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 218-2100  
; TELEFAX: (212) 218-2200  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3435  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; IMMEDIATE SOURCE:  
; LIBRARY: kidney  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 9..1655  
; IDENTIFICATION METHOD: E  
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-09-344-510B-7  
Query Match 8.3%; Score 43.4; DB 4; Length 3435;  
Best Local Similarity 53.9%; Pred. No. 1.7;  
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
QY 189 AGCTCGATCTCGTCCCGGCTCCACACACAGGGGAACTGCTCGGTGGCGAGCGGCAGC 248  
DB 574 ACCACGGGCTGACCGCCAGCTCCCGCAGCGCTCCACAGCGCGCGGTAGTAGCGCAGC 515  
QY 249 CCCAGCCGGGACAGCTCGTCCAGGCTGGCGTCCAGGACCTCTCGCGGTCCGGAGCGGAA 308  
DB 514 CCTCGCGGTTGGGACGCGCGCTGCCATTGGGGAGCACTCGCGCCACAGATGGAG 455  
QY 309 ACCGACAGGCTCAGATCCCTGTGCTAGTCGATCGATCGCTCAGTGGCG 353



db 454 AAGCGTATGTGAGTGAACCCCGAGCTCGGCGACGCGCTCCGTCTCG 410

Search completed: June 27, 2004, 19:41:58  
Job time : 96.8824 secs

Result No.	Query			Description		
	Score	Match	ID			
1	520	100.0	109519	12	US-09-758-759-1	Sequence 1, Appli
C 2	217	41.7	579	12	US-09-758-759-174	Sequence 174, App
C 3	59.6	11.5	15738	15	US-10-329-079-46	Sequence 46, Appl
C 4	59.6	11.5	6144	15	US-10-329-079-34	Sequence 34, Appl
5	55.2	10.6	2514	13	US-10-282-122A-28178	Sequence 28178, A
6	54.8	10.5	3833	13	US-10-087-192-707	Sequence 707, App
7	54.8	10.5	23833	13	US-10-087-192-706	Sequence 706, App
8	52.8	10.2	88421	9	US-09-976-059-1	Sequence 1, Appli
C 9	52.4	10.1	1059	16	US-10-369-493-24313	Sequence 24313, A
C 10	51.6	9.9	1293	13	US-10-329-027-17	Sequence 17, Appl
C 11	50.6	9.7	375	13	US-10-107-431-14	Sequence 14, Appl
C 12	50.6	9.7	45055	13	US-10-107-431-277	Sequence 277, App
C 13	50.4	9.7	661	16	US-10-260-238-3954	Sequence 3954, A
C 14	49.6	9.5	1326	15	US-10-156-761-5142	Sequence 5142, Ap

109060	DB	GGGTGACACAGGTCGACACAGGTTGGAGCACCAGCTCATCGCGCCTCGCGCCGGCA	109119
121	QY	TCGCGAAACACCGGGCCAGGATCACGTGCACACGCGCATCGCGCTCGATCTCGACGG	180
109120	DB	TGCGGAAACACCGGGCCAGGATCACGTGCACACGCGCATCGCGCTCGATCTCGACGG	109179
181	QY	TGGGCGGACAGTCGATCTCGTTCGCCCGGCTCCACACACGAGGAAACTGGCTCGGTGGCA	240
109180	DB	TGGGCGGACAGTCGATCTCGTTCGCCCGGCTCCACACACGAGGAAACTGGCTCGGTGGCA	109239
241	QY	GGGCGAGCCACGACCGGACAGCTCGTCCAGGCTGGCGGTCAACGGA	300
109240	DB	GGGCGAGCCACGACCGGACAGCTCGTCCAGGCTGGCGGTCAACGGA	109299
301	QY	GAGCGGAAACGCGCACGCTCAGATCCCTGTTCAGTCGCATCGGCTCAGTCCCGGTGCTCC	360
109300	DB	GAGCGGAAACGCGCACGCTCAGATCCCTGTTCAGTCGCATCGGCTCAGTCCCGGTGCTCC	109359
361	QY	CCCTTGGCTTGGAGGATAGCGGTTACGACGACGCGCACCAACGCGGGCGGGGGCGGG	420
109360	DB	CCCTTGGCTTGGAGGATAGCGGTTACGACGACGCGCACCAACGCGGGCGGGGGCGGG	109419
421	QY	GGGGTTACGCCGATCCGCTCGATGACACAGCGGCTTGGGGGTTCGGGGCGGTCCGCGAGATC	480
109420	DB	GGGGTTACGCCGATCCGCTCGATGACACAGCGGCTTGGGGGTTCGGGGCGGTCCGCGAGATC	109479
481	QY	CGTACCGCCCGACCGCTTCGGCCAGCGCCCGCGGATCC	520
109480	DB	CGTACCGCCCGACCGCTTCGGCCAGCGCCCGCGGATCC	109519

## RESULT 2

```

US-09-758-759-174/c
; Sequence 174, Application US/09758759
; Publication No. US20040101832A1
; GENERAL INFORMATION:
; APPLICANT: Hosted, Thomas J.
; APPLICANT: Wang, Tim X.
; APPLICANT: Horan, Ann C.
; TITLE OF INVENTION: Evernimicin Biosynthetic Genes
; FILE REFERENCE: ID0983K US
; CURRENT APPLICATION NUMBER: US/09/758,759
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/175,751
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 174
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Micromonospora carbonacea
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(579)
; OTHER INFORMATION: evmr2
US-09-758-759-174

```

```
Query Match      41.7%; Score 217; DB 12; Length 579;
Best Local Similarity 100.0%; Pred. No. 1.6e-44;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

1	GGTGACGACGAAACGACCGGTGTGTGCCCCCGCGCCCATCA	CGAACTGCAATCGCGCG	60
217	GGTGACGACGAAACGACCGGTGTGTGCCCCCGCGCCCATCA	CGAACTGCAATCGCGCG	158
61	GGGTGACCAAGTGTGACACCGGTGGGAGTTTGAGCAGCCAGCT	CAATGCGCGCTTGC	120
157	GGGTGACCAAGTGTGACACCGGTGGGAGTTTGAGCAGCCAGCT	CAATGCGCGCTTGC	98
121	TGCGGAAAACCGCGGCCAGGATCAGTGTGACACGCGCGAT	TGCGCGCTCGATTCGACGG	180
97	TGCGGAAAACCGCGGCCAGGATCAGTGTGACACGCGCGAT	TGCGCGCTCGATTCGACGG	38

```

Qy      181  TCGCCCGCAGCTCGATCTCGTCGCCCGGCTCCACAC 217
Db      37   TCGCCCGCAGCTCGATCTCGTCGCCCGGCTCCACAC 1

RESULT 3
US-10-329-079-46/c
/ Sequence 46, Application US/10329079
/ Publication No. US2003019891A1
/ GENERAL INFORMATION:
/ APPLICANT: FARNET, Chris
/ APPLICANT: ZAZOPOULOS, Emmanuel
/ APPLICANT: STAPPA, Alfredo
/ TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN T
/ FILE REFERENCE: 3002-11US
/ CURRENT APPLICATION NUMBER: US/10/329,079
/ CURRENT FILING DATE: 2002-12-24
/ NUMBER OF SEQ ID NOS: 66
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 46
/ LENGTH: 15738
/ TYPE: DNA
/ ORGANISM: Streptomyces refluens
US-10-329-079-46

```

Query Match	11.5%;	Score 59.6;	DB 15;	Length 15738;
Best Local Similarity	46.6%;	Pred. No. 7.1e-06;		
Matches 191;	Conservative 0;	Mismatches 219;	Indels 0;	Gaps 0;
QY	1	GGTGACGACGAAACACACCGGTGGTGGCCCCCGCGCGCCATCAGAACTCCCACTCCGGCGG	60	
DB	7879	GCTGCTGGCGGTAGGAGACGGCAGCCGCGCGCGCGGCACACCGCCGCGCACCGCGC	7820	
QY	61	GGGTGACCAAGGTGCACCAAGTGGGAGTTGAGCAGCAGCTCATCGCGCTGGCGCCGGCA	120	
DB	7819	CGGGCGCGGCTTCGGGGCGGTTCGAGGTGCGCGGCCAGCCCGGCACCGTGGGGCGCTCGA	7760	
QY	121	TGCCGAAACACCGCGCCAGGATCAGTGCAGCAGCCGCGATGCGCGCTCGATCTCGACGG	180	
DB	7759	AGAGGGCGTCGATGGCGGAGTCCAGTTCGAGCAGGGCCCTGATCGGGTGTATCACACCA	7700	
QY	181	TGGCGCGCAGCTCGATCTCGTGGCGCGGCTCCACACACGAGGGGAACTGGCTCGGTGGCA	240	
DB	7699	TCGCCAGCAGGAGTGGCGGCCCGCAGGGCGAAGAAGTGTGCTCCACCCGACCCCGGCA	7640	
QY	241	CGGCGAGCCCGACGCGGGAACAGCTCGTTCAGAGCTGGCGTCAACGACCTTCGGGGGTGG	300	
DB	7639	CCCCAGCACTTCGGCGAAACAACCGCACAGGATCTCTTCGCGCGACCGCGCGCGCCC	7580	
QY	301	GAGCGGAAACGGCGCAGGCTCAGATCCCTGTTCAGTTCGATCGGCTCAGTTCGCGGTCTCC	360	
DB	7579	GCCCCACCCCGGCACCGGGCGGAATCCGGCACCGGCAACGGGGCCCGGTCCAGTTTC	7520	
QY	361	CCCTTGGCTGGGATAGCGTTTCACGACGAGCGGCACCAACGGCGGC	410	
DB	7519	CGTTGGCGCTCAACGCAACCCGCTCTCCAGACCCACCAACGCGCCACGCGC	7470	

RESIST 4

```

US/10-329-079-34/c
; Sequence 34, Application No./10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: ZAZOPOULOS, Emmanuel
; APPLICANT: STAFFA, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS IN
; FILE REFERENCE: 3002-11US
; CURRENT APPLICATION NUMBER: US/10/329,079
; CURRENT FILING DATE: 2002-12-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34

```

; LENGTH: 61944;  
; TYPE: DNA;  
; ORGANISM: Streptomyces refuineus  
US-10-329-079-34

Query Match 11.5%; Score 59.6; DB 15; Length 61944;  
Best Local Similarity 46.6%; Pred. No. 5.8e-06;  
Matches 191; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

QY 1 GGTGACAGCAGCAACGACCGGTGCTGCGCCCGCGCCCATCAGCAACTGCCACTCCGCGC 60  
DB 38215 GCTGCTGGCGTAGAGAGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38156

QY 61 GGGTGACACAGGTGACACAGTGGAGTGGAGCAGCAGCTATCCGCCCTGCGCGCGCA 120  
DB 38155 CGCGCGCGCGCTGCGCGCGCGTGGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCTCGA 38096

QY 121 TCCGAAACACCGCGCCAGGATCAGTGACAGCAGCGCGATCGCGCGCGCGCGCGCGCGCG 180  
DB 38095 AGAGGGCTGATGCGCGCGCGTCCAGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 38036

QY 181 TCGCGCGCAGCTGATCTGTCGCGCGCGCGTCCACACAGCGGGAACCTGGCTCGGTGGCA 240  
DB 38035 TCGCGCAGAGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37976

QY 241 GCGGAGCGCCAGCGCGGACAGCTGTCAGGCTGGCGCTCAGGACCTCTCGCGGTGG 300  
DB 37975 CCCCAGACCTTCGCGGAAACACCGCGCAGAGATCTCTCGCGCGCGCGCGCGCGCG 37916

QY 301 GAGCGGAAACGCGCGCGCTGATCTGTCGCGCGCGCGTCCAGTGGATCGGTTCAGTGGCTGCTGCC 360  
DB 37915 GCGCCACCG 37856

QY 361 CCTTGGCTGGAGAGTAGGGTTCAGCAGAGCGCGGACACACGCGCGCGCGCGCGCG 410  
DB 37855 CTTTGGCGGTCAACGGCAACCGCTCTCAGCAGCAGCGCGCGCGCGCGCGCGCGCG 37806

## RESULT 5

US-10-282-122A-28178  
; Sequence 28178, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangshu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELTRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 28178  
; LENGTH: 2514  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-282-122A-28178

Query Match 10.6%; Score 55.2; DB 13; Length 2514;  
Best Local Similarity 47.5%; Pred. No. 0.00011;  
Matches 194; Conservative 0; Mismatches 213; Indels 1; Gaps 1;

QY 109 CTCTGCGCGCGATGCGCGGCAACACCGCGCGCGAGGATCAGTGCAGCAGCGGATCGCGCCT 168  
DB 242 CTCTCGGTGCGCGCGGTATCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 301

QY 169 CGATCTCGAGGTCGCGCGCGCGCTCGATCTCGTTCGCGCGCGCGCGCGCGCGCGCGCG 228  
DB 302 CGATCAACAGATTCTTCTGCGCGAATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361

QY 229 GGCTCGGTGCG 288  
DB 362 CG 421

QY 289 CTGCGCGGTGCG 348  
DB 422 CGGAGGATCCG 481

QY 349 TGGCGGTGCTGCG 408  
DB 482 GGTCTCATCG 541

QY 409 GCGCGCGCGCGCGCGCGGTTCAGCGCGATCCGCTCGATCAGCAGCGCGGTTCGCGCGCG 468  
DB 542 GCG 600

QY 469 GTGCGCGGATCGGTACCG 516  
DB 601 CCGCGCGGATTCGACGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648

## RESULT 6

US-10-087-192-707  
; Sequence 707, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 707  
; LENGTH: 3833  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-087-192-707

Query Match 10.5%; Score 54.8; DB 13; Length 3833;

Best Local Similarity 46.2%; Pred. No. 0.00013;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
QY 46 ACTGCATCTCCGCGCGGTGACCAAGTTCGACAGAGTGGAGATTGAGCAGCCAGCTCATCG 105  
Db 758 AGTTCCGCGCGGCAAGTGAAGCAATCTGCGGGTCTGTCGGCCACGACCCCATCAGCG 817  
QY 106 CGCGCTGCGCGCGGATGCGAAACACCGCGGCCAGGATCAGCTGACGACAGCGCGATGCGCG 165  
Db 818 TCCCGCGCGCGCGCCACCACTGCTGCTCACTGCTCCCGCCAGCAGAGCGGCTGCG 877  
QY 166 CTTGATCTGACAGTTCGCGCGGAGTCTGATCTGCTGCGCGGCTTCCACACAGGGGAA 225  
Db 878 TCGCGCGCTGCGGCAACGACAGCTGCGCGGCGACCCCGCGCGCGCCAAAGTGCAGCTCGT 937  
QY 226 ACTGCTCGGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 285  
Db 938 GCCGGTGGCGACCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 997  
QY 286 CTTTCGCGGCTGCGGAGCGGAAACGCGCATCGCTGCGCGGCTGCGCGGCTGCGCGGCT 345  
Db 998 GCGCGAGCGCGGTGTCGGTGGCTGGAGC-CCCTGAGCGCGCTCCATGTCCTGCGCGCGCC 1056  
QY 346 CAGTCCCGCTGCTCCCTTGGCGTGGAGGATGCGGTTACAGAGCGGCGGCGGCGGCGG 405  
Db 1057 CTGAGCGCTGCGACACCTGAGCGCGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 1116  
QY 406 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 465  
Db 1117 CTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1176  
QY 466 GCGGTGCGGAGATCCGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 515  
Db 1177 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1226

RESULT 8  
US-09-976-059-1  
; Sequence 1, Application US/09976059  
; Patent No. US2002016474A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Zazopoulos, Emmanuel  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
; FILE REFERENCE: 3019-PC  
; CURRENT APPLICATION NUMBER: US/09/976,059  
; CURRENT FILING DATE: 2001-10-15  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 88421  
; TYPE: DNA  
; ORGANISM: Actinoplanes sp.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2077)..(3078)  
; OTHER INFORMATION: ORF 1; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (3118)..(4032)  
; OTHER INFORMATION: ORF 2; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (4038)..(5048)  
; OTHER INFORMATION: ORF 3; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (4038)..(5048)  
; OTHER INFORMATION: ORF 3; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (6665)..(5814)  
; OTHER INFORMATION: ORF 4; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (7703)..(6693)  
; OTHER INFORMATION: ORF 5; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (9464)..(8130)  
; OTHER INFORMATION: ORF 6; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (9691)..(10761)  
; OTHER INFORMATION: ORF 7; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (12751)..(10829)

Best Local Similarity 46.2%; Pred. No. 0.00013;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
QY 46 ACTGCATCTCCGCGCGGTGACCAAGTTCGACAGAGTGGAGATTGAGCAGCCAGCTCATCG 105  
Db 758 AGTTCCGCGCGGCAAGTGAAGCAATCTGCGGGTCTGTCGGCCACGACCCCATCAGCG 817  
QY 106 CGCGCTGCGCGCGGATGCGAAACACCGCGGCCAGGATCAGCTGACGACAGCGCGATGCGCG 165  
Db 818 TCCCGCGCGCGCGCCACCACTGCTGCTCACTGCTCCCGCCAGCAGAGCGGCTGCG 877  
QY 166 CTTGATCTGACAGTTCGCGCGGAGTCTGATCTGCTGCGCGGCTTCCACACAGGGGAA 225  
Db 878 TCGCGCGCTGCGGCAACGACAGCTGCGCGGCGACCCCGCGCGCGCCAAAGTGCAGCTCGT 937  
QY 226 ACTGCTCGGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 285  
Db 938 GCCGGTGGCGACCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 997  
QY 286 CTTTCGCGGCTGCGGAGCGGAAACGCGCATCGCTGCGCGGCTGCGCGGCTGCGCGGCT 345  
Db 998 GCGCGAGCGCGGTGTCGGTGGCTGGAGC-CCCTGAGCGCGCTCCATGTCCTGCGCGCGCC 1056  
QY 346 CAGTCCCGCTGCTCCCTTGGCGTGGAGGATGCGGTTACAGAGCGGCGGCGGCGGCGG 405  
Db 1057 CTGAGCGCTGCGACACCTGAGCGCGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 1116  
QY 406 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 465  
Db 1117 CTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1176  
QY 466 GCGGTGCGGAGATCCGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 515  
Db 1177 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1226

RESULT 7  
US-10-087-192-706  
; Sequence 706, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 706  
; LENGTH: 23833  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(23833)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-706  
Query Match 10.5%; Score 54.8; DB 13; Length 23833;  
Best Local Similarity 46.2%; Pred. No. 0.0001;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
QY 46 ACTGCATCTCCGCGCGGTGACCAAGTTCGACAGAGTGGAGATTGAGCAGCCAGCTCATCG 105  
Db 10758 AGTTCCGCGCGGCAAGTGAAGCAATCTGCGGGTCTGTCGGCCACGACCCCATCAGCG 10817  
QY 106 CGCGCTGCGCGCGGATGCGAAACACCGCGGCCAGGATCAGCTGACGACGCGCGATGCGCG 165

```
OTHER INFORMATION: ORF 8; negative strandedness
NAME/KEY: misc feature
LOCATION: (13617)..(12802)
OTHER INFORMATION: ORF 9; negative strandedness
NAME/KEY: misc feature
LOCATION: (15203)..(13614)
OTHER INFORMATION: ORF 10; negative strandedness
LOCATION: (15591)..(15863)
OTHER INFORMATION: ORF 11; positive strandedness
NAME/KEY: misc feature
LOCATION: (15880)..(19035)
OTHER INFORMATION: ORF 12; positive strandedness
NAME/KEY: misc feature
LOCATION: (19032)..(39713)
OTHER INFORMATION: ORF 13; positive strandedness
NAME/KEY: misc feature
LOCATION: (39713)..(58900)
OTHER INFORMATION: ORF 14; positive strandedness
NAME/KEY: misc feature
LOCATION: (65826)..(66530)
OTHER INFORMATION: ORF 15; positive strandedness
NAME/KEY: misc feature
LOCATION: (66546)..(67370)
OTHER INFORMATION: ORF 16; positive strandedness
NAME/KEY: misc feature
LOCATION: (67384)..(70059)
OTHER INFORMATION: ORF 17; positive strandedness
NAME/KEY: misc feature
LOCATION: (70059)..(70662)
OTHER INFORMATION: ORF 18; positive strandedness
NAME/KEY: misc feature
LOCATION: (70659)..(71906)
OTHER INFORMATION: ORF 19; positive strandedness
NAME/KEY: misc feature
LOCATION: (73439)..(71964)
OTHER INFORMATION: ORF 20; negative strandedness
NAME/KEY: misc feature
LOCATION: (74216)..(73563)
OTHER INFORMATION: ORF 21; negative strandedness
NAME/KEY: misc feature
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
NAME/KEY: misc feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
NAME/KEY: misc feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
NAME/KEY: misc feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
NAME/KEY: misc feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
NAME/KEY: misc feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
NAME/KEY: misc feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
NAME/KEY: misc feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
NAME/KEY: misc feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
NAME/KEY: misc feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
NAME/KEY: misc feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; positive strandedness
```

```
NAME/KEY: misc feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-758-059-1

Query Match      10.2%; Score 52.8; DB 9; Length 88421;
Best Local Similarity 46.5%; Pred. No. 0.00026;
Matches 240; Conservative 0; Mismatches 272; Indels 4; Gaps 2;

QY      7 GCACGAAACGACGGTGGTCCCGCCCGCCGATCAAGAACTGCACACTCGCGCGGGTGA 66
DB      67518 GCACGAGACGACGTGGCGCGCCCGCGCGCGATGAGCGCGCGCTCGCGCGCGCGCGA 67577
QY      67 CCAGGTCGACACGAGTGGAGTTGAGCAGCCAGCTCATCGCCGCTGCGCGCGCATCCGA 126
DB      67578 ACTCGCCCGCCCGGTGTCGCCCGGAGCCCGGTGCGGGGGGTGCTGCTCGCTACCGGA 67637
QY      127 AACACCGCGCCAGGATCACGTGCGACGACGCGGCGATGCGCGCTCGATCTCGACGGTCCGCC 186
DB      67638 CGGCTCGCGCGACCTGGTGTGTCGCCCGCGCGCGCGCTGGACCGCGCGCGCTGAT 67697
QY      187 GCAGCTCGATCTGTCGCCCGCGCTCCACACAGGGGAACTGGTCTGGTGGCAGCGCGCA 246
DB      67698 CGCCCTGGCGCCCGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67757
QY      247 GCGCCAGCGCGGACAGCTGTCAGGCTGGCGCTACGCGACCTCTCGCGGGTTCGGGAGCGG 306
DB      67758 GCGCGCCTCGCGCGCGCGCGCTGGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCG 67817
QY      307 AAAGCGCGAGGCTCAGATCCCTGTGTCAGTCGATCGGCTCAGTCCGGTTCGCTCCCTTG 366
DB      67818 CGAGCTGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 67875
QY      367 GCCTGGGAGGATAGCGGTTTCAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 424
DB      67876 GCGAAGTCTCTCGCGCGGTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67935
QY      425 TTACGCGGATCCGCTCGATGACAGCGGCTGCGGGGTTCGGGGGTTCGGCGGAGATCCGTA 484
DB      67936 TTGGGCGCGCGCGCGCGCGCTCGGCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 67995
QY      485 CCGCCCGGACCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 520
DB      67996 GCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 68031

RESULT 9
US-10-369-493-24313/c
Sequence 24313, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 24313
LENGTH: 1059
TYPE: DNA
ORGANISM: Deinococcus radiodurans
US-10-369-493-24313

Query Match      10.1%; Score 52.4; DB 16; Length 1059;
Best Local Similarity 48.6%; Pred. No. 0.00063;
Matches 143; Conservative 0; Mismatches 151; Indels 0; Gaps 0;
```

QY 23 GTCCGCCGCGCCGATCAGAACTGCACTCGCGCGGGGTGACCAAGTTCGACCAAGTG 82  
Db 795 GTCCGCCGCGGGTGGCCAGGGCATCAGGCCGAGGACCAAGGAAATCAGCGCGCGGAGTC 736  
QY 83 GGAGTTGAGCAGCAGCTCATCGCGGCTCGCGCGCATCGCGGAAACACCGCGCCAGGAT 142  
Db 735 GGAAAGGATGTCGCGAAGAGTTCTCGGTGAGATCACTGTCGAGCGGTGGGTGCGC 676  
QY 143 CAGTGTGAGCAGCGGATGCGGCTGATCTGACGCTGCGCGCGAGCTCGATCTCGTC 202  
Db 675 CACGATCAGCATCGCCAGCTGTCCAGTACTCGTGTTCAGGTGAATCAGCGGTAGTC 616  
QY 203 GCCCGGCTCCACACCAAGGGAACCTGGCTCGGTGCGAGCGGAGCCAGCCGCGGACAG 262  
Db 615 GCGTTCAGCAAGCGCTGACGTGCGCGCCCACTCGGACATTCGAGCAGCTTCG 556  
QY 263 CTGCTCAGGCTGCGTCAAGGACTCTCGGGGTGCGGAGCGGAACGCGCAC 316  
Db 555 CTTGTCCAGCTGTGTCAGCGGCCCTTGCCTGCTCGCGCGCCCAAGAGCGAC 502

RESULT 10  
US-10-329-027-17/C  
; Sequence 17, Application US/10329027  
; Publication No. US20030211567A1  
; GENERAL INFORMATION:  
; APPLICANT: STAFFA, Chris  
; APPLICANT: ZAZOPOULOS, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS, METHODS AND SYSTEMS FOR DISCOVERY OF LIPOPEPTIDES  
; FILE REFERENCE: 8822/2002  
; CURRENT APPLICATION NUMBER: US/10/329, 027  
; CURRENT FILING DATE: 2002-10-24  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 17  
; LENGTH: 1293  
; TYPE: DNA  
; ORGANISM: Actinomyces

Query Match 9.9%; Score 51.6; DB 13; Length 1293;  
Best Local Similarity 48.0%; Pred. No. 0.00096;  
Matches 212; Conservative 0; Mismatches 224; Indels 6; Gaps 2;

QY 23 GTCCGCCGCGCCGATCAGAACTGCACTCGCGCGGGGTGACCAAGTTCGACCAAGTG 82  
Db 522 GTCCGCCCTGCTCAGCGCTCGGAAGACCCGCTCGCGCGGGGTCCGGCGCGCCAGGGC 463  
QY 83 GGAGTTGAGCAGCAGCTCATCGCGGCTCGCGCGCATGCCGAAACACCGCGCCAGGAT 142  
Db 462 GCGTAGACCTGCGCACCCCGGGGGGTGAGACGATGCGCGAAGCGTTCAGCAGCGC 403  
QY 143 CAGTGTGAGCAGCGGATGCGGCTGATCTGACGCTGCGCGCGAGCTCGATCTCGTC 202  
Db 402 GTGTGTGAGCGGTGGAACACACAGCCGCGGCTCGCGCGGAGCAGCAGCGCTGGCG 343  
QY 203 GCCCGGCTCCACACCAAGGGAACCTGCTCGGTGCGAGCGGAGCCCGCGGAGCAG 262  
Db 342 GCTCAGCGGGCCCGGCGCAGGCGGAGACTCGGCCATTCGGGCCCCCATTCAGCGCCG 283  
QY 263 CTGCTCAGGCTGCGGTCAAGGACTCTGCGGGTTCGGGAGCGGAACGCGCACGG --C 319  
Db 282 GCGCGCGGCTCCGGGTGCGGCTCGCGCGCAGGTGCGACACTCCAGCGGCTCGGGTC 223  
QY 320 TCAGATCCCTGATGTCATCGGCTCAGTCCGCTGCTCCCTCTGCGCTGGAGATA 379  
Db 222 GCGCGCGCGGACGAGCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 163  
QY 380 CCGGTTACGACGAGCG ---GCACACGCGCGGCGGGGGCGGGGCGGTTTCAGCCGATCC 436  
Db 162 CTCGCTCTCGCGCAGCGGTGTCGCGACCGCGCGGTTCGAGCAGCCCGGGATCGCGCGGCC 103

QY 437 GCTCGATGACCAAGCGCGGCTGCGG 458  
Db 102 GATCTCGTAGTAGCAGCGCGCAG 81

RESULT 11  
US-10-107-431-14  
; Sequence 14, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: FARNET, Chris  
; APPLICANT: STAFFA, Alfredo  
; APPLICANT: ZAZOPOULOS, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 14  
; LENGTH: 975  
; TYPE: DNA  
; ORGANISM: Streptomyces mobaraensis

US-10-107-431-14  
Query Match 9.7%; Score 50.6; DB 13; Length 975;  
Best Local Similarity 50.9%; Pred. No. 0.0018;  
Matches 145; Conservative 0; Mismatches 139; Indels 1; Gaps 1;

QY 23 GTCCGCCGCGCGCCATCAGAACTGCACTCGCGCGGGGTGACCAAGTTCGACCAAGTG 82  
Db 402 GCGCGCGCGCGCGAGCAGCACTAGCGCAAGCCAGCTCGCGGAGCCCGCGCGTCT 461  
QY 83 GGAGTTGAGCAGCAGCTCATCGCGGCTCGCGCGCATGCCGAAACACACCGCGGCGAGAT 142  
Db 462 GGAGCGCGCGCGGAGGGGTGCTCGAGCGGGGTGCTCGGATCGCAATGTCGCGG 521  
QY 143 CAGTGTGAGCAGCGGAGTGGC -CGCTCGATCTGACGCTGCGCGCGAGCTCGATCTCT 201  
Db 522 CCGCGGACCCCGCGCGGTGAGCTGCTGGCGCGGTGCGCGCGGCTCGCGGAGCGGT 581  
QY 202 CCGCGCGCTCCACACCAAGGGAACCTGGCTGGTGGCGGCGAGCGCGCGCGGAGCA 261  
Db 582 CCGCGGAGCACCTCCCGCGGTGCTGAGCTGTCCAGCTGCGCGCGGAGCGGACTA 641  
QY 262 GCTGTCCAGGTGCGGTGACGAGCTTCGCGGGTTCGGAGCGG 306  
Db 642 CGTCGACGTGCGCAGCTGTGGAAGCGGTGCTGCGCGCGCCACCG 686

RESULT 12  
US-10-107-431-277  
; Sequence 277, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: FARNET, Chris  
; APPLICANT: STAFFA, Alfredo  
; APPLICANT: ZAZOPOULOS, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 277  
; LENGTH: 45055  
; TYPE: DNA  
; ORGANISM: Streptomyces mobaraensis

US-10-107-431-277

Query Match: 9.7%; Score 50.6; DB 13; Length 45055;  
Best Local Similarity 50.9%; Pred. No. 0.001;  
Matches 145; Conservative 0; Mismatches 139; Indels 1; Gaps 1;  
QY 23 GTGCCCCCGCGCCCATACGAACCTGCCACTCCGGCGGGGTGACCAAGTTCAGCAGGTG 82  
Db 44482 GCGGGCGCGCGACGACCACTAGCGCAAGCCCAAGCTCGCGGAGCGCGCGTGTCT 44541  
QY 83 GGAGTTGACAGCAGCTCATCGCGCTCGCGCGGATGCGCGGATGCGGAACACCGGCGCAGGAT 142  
Db 44542 GGAGGCGCGCGGAGGGGCTGTGAGCGGGCGCTGTGCGGATGCGGAATGTGCGGG 44601  
QY 143 CAGCTGACAGCGCGCATCGC-CGCTCGATCTCAGCGTCCGCGCGCAGCTCGATCTGT 201  
Db 44602 CCGGGACACCGCGGCTCAGCTGTGCGCGGGTGGCGGCGGCTCGCGAGCGCGT 44661  
QY 202 CGCCCGGCTCCACACAGCGGGAACCTGCTCGTGGGAGCGGAGCGGAGCGGAGCA 261  
Db 44662 CGCCCGGAGCACCTCCCGCGGCTGTGAGGCTGTCCAGCTGCGCGGCCACCGGAGTA 44721  
QY 262 GCTGCTCAGGCTGGGCTCAGGACCTCTCGCGGGTTCGGGAGCGG 306  
Db 44722 CGTGGAGCTGCGGAGCTGTGCGAGCGTGTGCGCGGACCG 44766

RESULT 13  
US-10-260-238-3954/c  
; Sequence 3954, Application US/10260238  
; Publication No. US20040016025A1  
; GENERAL INFORMATION:  
; APPLICANT: Budworth, Paul R.  
; APPLICANT: Moughamer, Todd G.  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiyuki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricke, Darrell  
; APPLICANT: Zhu, Tong  
; FILE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
; FILE REFERENCE: 60111-NP  
; CURRENT APPLICATION NUMBER: US/10/260,238  
; PRIOR APPLICATION NUMBER: US 60/325,448  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; NUMBER OF SEQ ID NOS: 6077  
; SEQ ID NO 3954  
; LENGTH: 661  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
US-10-260-238-3954

Query Match: 9.7%; Score 50.4; DB 16; Length 661;  
Best Local Similarity 47.5%; Pred. No. 0.0021;  
Matches 150; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
QY 47 CTGCCACTCCGCGGGGTGACCAAGTTCAGGTCAGGAGTTCAGCAGCAGCTCATCGC 106  
Db 457 CCGCGTCTGTCAGGGCATCGGAGCGCGGCGCGGCGGAGCGGCGGCGCGCGCGTGC 398  
QY 107 CGCTTGGCGCGGATGCGGAACACCGGCGCAGGATCAGTGCAGCAGCGGCGATCGCGC 166  
Db 397 CGCGGTCTGCGGATGCTCAGCTCTCCACACCGGCTCGCGGCTCTGCCCATGAGCGC 338  
QY 167 CTCGATCTCAGCGTGGCGCGGAGCTCGATCTCGTGGCGGCTCCACCGGAGAA 226  
Db 337 CACCAGCGCTCGGGGCTCGGCTCTGCTCTCGAACCCTCTCCGCCACCGGCCCTC 278

QY 227 CTGCTCGTGGCAGCGGACGCCCCAGCCGCGGACAGCTCTCCAGGTGGCGGTCAACGGAC 286  
Db 277 CCGCGCGGTGCGAGAAGCCCGCTTCAGGCTCGCAGGTCTCTCCACCATGCGGTGTC 218  
QY 287 CTCTCGGGGTGGGAGCGGAAACCGGACGGCTCAGATCTCTGTCTAGTCTGATCGGCTC 346  
Db 217 CTCCCTCGTGAAGCTCGCTGGTGTGCGCGGCGAGCAGCACCATCAGGAACGCTGGAA 158  
QY 347 AGTCCCGGTCTGTCGCC 362  
Db 157 CGAGGCTTCATCACC 142

RESULT 14  
US-10-156-761-5142/c  
; Sequence 5142, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 5142  
; LENGTH: 1326  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1326)  
US-10-156-761-5142

Query Match: 9.5%; Score 49.6; DB 15; Length 1326;  
Best Local Similarity 48.9%; Pred. No. 0.003;  
Matches 133; Conservative 0; Mismatches 139; Indels 0; Gaps 0;  
QY 43 CGAAGTCCCACTCGCGCGGGTCAACAGGTTCGACCGAGTGGAGTTCAGCAGCAGTCA 102  
Db 553 CGAGTACGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 494  
QY 103 TCGCGGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 162  
Db 493 TCTGATCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 434  
QY 163 GCGCTCGATCTCGAGGTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 222  
Db 433 CCAGTCTGTTGAGGATGACAGGATCAGCGTCTCCAGGAGCAGCAGCTCCGCGAAG 374  
QY 223 GAAACTGGCTTCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 314  
Db 373 AGCTCTCGCGGTCGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 314  
QY 283 GGACTCTTCGCGGTTCGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 282  
Db 313 AGATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 282

RESULT 15  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1



```

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match          9.5%; Score 49.6; DB 15; Length 9025608;
Best Local Similarity 48.9%; Pred. No. 0.00082;
Matches 133; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 43 CGAAGTGCACCTCGCGGGGTGACAGAGTGCAGAGTGGAGTTGAGCAGCGCCAGCTCA 102
Db 6266553 CGAGTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6266612

QY 103 TCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 162
Db 6266613 TCTGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6266672

QY 163 CGCGCTTCATCTCGACCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 222
Db 6266673 CGAGTCTGTTGAGAGTGCAGAGGATCAGCGTCTCCAGGAGCAGCACTCCG 6266732

QY 223 GAAGTGCCTGTTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 282
Db 6266733 AGCCCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6266792

QY 283 CGACCTCTCGCGGGTCCGCGAGCGGAAACGCGC 314
Db 6266793 AGATGTCCTCCCGCGAGCGGTACGAGCGGAC 6266824

RESULT 16
US-10-156-761-4136/c
; Sequence 4136, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match          9.5%; Score 49.6; DB 15; Length 9025608;
Best Local Similarity 48.9%; Pred. No. 0.00082;
Matches 133; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 43 CGAAGTGCACCTCGCGGGGTGACAGAGTGCAGAGTGGAGTTGAGCAGCGCCAGCTCA 102
Db 6266553 CGAGTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6266612

QY 103 TCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 162
Db 6266613 TCTGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6266672

QY 163 CGCGCTTCATCTCGACCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 222
Db 6266673 CGAGTCTGTTGAGAGTGCAGAGGATCAGCGTCTCCAGGAGCAGCACTCCG 6266732

QY 223 GAAGTGCCTGTTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 282
Db 6266733 AGCCCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6266792

QY 283 CGACCTCTCGCGGGTCCGCGAGCGGAAACGCGC 314
Db 6266793 AGATGTCCTCCCGCGAGCGGTACGAGCGGAC 6266824
```

```

; SEQ ID NO 4136
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1596)
US-10-156-761-4136

Query Match          9.5%; Score 49.4; DB 15; Length 1596;
Best Local Similarity 47.8%; Pred. No. 0.0032;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 2 GTGCAGCAGGAACGACCGGTGTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 61
Db 1215 GAGACGCTCGTAGAGTGGCGTGCCTCGCTCAGCCACAGGTGCGACCGCGCGAT 1156

QY 62 GGTGACCAAGTCGACCAAGTGGGAGTTGAGCAGCGAGCTCATCGCGCGCTTGGCGCG 121
Db 1155 GGCTAGCTGTGCGCGGTCCAGTGTGGTTCAGTCTGTCACCAAGTTGCGCTCGCG 1096

QY 122 GCCGAAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 181
Db 1095 GACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1036

QY 182 CGCGCGCGAGTCGATCTCGTTCGCCCGGTCCACACAGCGGGAACCTGCGTGGCG 241
Db 1035 CACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 976

QY 242 CGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 975 CGCGCGTTCCAGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 917

RESULT 17
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match          9.5%; Score 49.4; DB 15; Length 9025608;
Best Local Similarity 47.8%; Pred. No. 0.00092;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 2 GTGCAGCAGGAACGACCGGTGTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 61
Db 5099716 GAGACGCTCGTAGAGTGGCGTGCCTCGCTCAGCCACAGGTGCGACCGCG 5099657
```

QY 62 GGTGACAGAGTGCACACAGGTGGAGTTGAGACAGCCAGTCTATCGCGCCCTGCGCCGGCAT 121  
Db 5099656 GGTGACAGTCTGCGCCGCTCCAGTGGTGGGTGAGTCTGTCACAGGTTGCGGTTCGCGGCC 5099597  
QY 122 GCCGAAACACCGCGCCAGAGTACAGTGCAGACGGGATGCGGCTTCGATCTCGACGCT 181  
Db 5099596 GACCGGGACACCCAGCAGTCTGCGCCCTCGGCACAGGGGACAGGACTGCGTCTCAGCGC 5099537  
QY 182 CGCGCAGTCTGATCTGTCGCGCGGCTCCACACAGGGGAACTGGCTCGGTGGCAG 241  
Db 5099536 CACCGCAGTCTGCTGTCGCGGACACAGCGCTGAGCGGTGAAAGGTACGGGCGGAG 5099477  
QY 242 CGCGAGCCCGCAGCGGACAGTCTGTCAGGCTGGGCTCAGCGACCTCTCGCGGGTGG 300  
Db 5099476 CCGCGCTTCAGCGGCGGAGTGGCGGGGTGACGGAACGGTACGCTCGGTGTCG 5099418

## RESULT 18

US-10-282-122A-25414/c

; Sequence 25414, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELIUPA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 25414

; LENGTH: 1575

; TYPE: DNA

; ORGANISM: Mycobacterium avium

US-10-282-122A-25414

## Query Match

; Sequence 25414, Score 49.2; DB 13; Length 1575;

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELIUPA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 25414

; LENGTH: 1575

; TYPE: DNA

; ORGANISM: Mycobacterium avium

US-10-282-122A-25414

Query Match

; Sequence 25414, Score 49.2; DB 13; Length 1575;

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELIUPA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 25414

; LENGTH: 1575

; TYPE: DNA

; ORGANISM: Mycobacterium avium

US-10-282-122A-25414

Query Match

; Sequence 25414, Score 49.2; DB 13; Length 1575;

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELIUPA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 25414

; LENGTH: 1575

; TYPE: DNA

; ORGANISM: Mycobacterium avium

US-10-282-122A-25414

Query Match

; Sequence 25414, Score 49.2; DB 13; Length 1575;

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELIUPA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 25414

; LENGTH: 1575

; TYPE: DNA

; ORGANISM: Mycobacterium avium

US-10-282-122A-25414

Query Match

; Sequence 25414, Score 49.2; DB 13; Length 1575;

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELIUPA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242

APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 39530  
LENGTH: 711  
TYPE: DNA  
ORGANISM: Xanthomonas campestris  
US-10-369-493-39530

Query Match 9.4%; Score 49; DB 16; Length 711;  
Best Local Similarity 52.2%; Pred. No. 0.0046;  
Matches 109; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 154 CGGGATCGCGCTCGATCTCGAGCTGGCGCGACCTGATCTCGTCCCGGCTCCC 213  
DB 406 CGGGGCGCGCTCCACCACTTCGGTGGAGACCGCGCCAGCTGAAATTGGACGCC 347  
QY 214 ACACAGGGGAACTGGCTCGGTGGCAGCGGACCGCCCGGACAGCTCGTCCAGGC 273  
DB 346 ACACAGCGGAATCTGGCCGCTGCTTGCAGCAAGCGGCCGCTGCGCTCGTCCAGCC 287  
QY 274 TGGCGTCACGACCTCTCGCGGCTGGGAGCGGAAACGCGACGGCTCAGATCCCTGTCA 333  
DB 286 CGGTGGTCCCGAGACAGCGGTTTGGCGGCTGCACGCAACAGCCAGGATCGGGCAA 227  
QY 334 GTCCGATCGGCTCAGTGGCGGCTCGCTCCC 362  
DB 226 AGCCTGGCGGAGGCTGAAGTCGATCGCC 198

RESULT 21  
US-10-369-493-39909/c  
Sequence 39909, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 39509  
LENGTH: 711  
TYPE: DNA  
ORGANISM: Xanthomonas campestris  
US-10-369-493-39909

Query Match 9.4%; Score 49; DB 16; Length 711;  
Best Local Similarity 52.2%; Pred. No. 0.0046;  
Matches 109; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 154 CGGGATCGCGCTCGATCTCGAGCTGGCGCGACCTGATCTCGTCCCGGCTCCC 213  
DB 406 CGGGGCGCGCTCCACCACTTCGGTGGAGACCGCGCCAGGCTGAAATTGGACGCC 347  
QY 214 ACACAGGGGAACTGGCTCGGTGGCAGCGGACCGCCCGGACAGCTCGTCCAGGC 273  
DB 346 ACACAGCGGAATCTGGCCGCTGCTTGCAGCAAGCGGCCGCTGCGCTCGTCCAGCC 287  
QY 274 TGGCGTCACGACCTCTCGCGGCTGGGAGCGGAAACGCGCAGCGCTCAGATCCCTGTCA 333

DB 286 CGGTGGTCCCGAGACCAAGCGGTTTCCCGCTGCACGCAACAGCCAGGATCGGCGAA 227  
QY 334 GTCCGATCGGCTCAGTGGCGGCTCGTCCC 362  
DB 226 AGCCTGGCGGAGGCTGAAGTCGATCGCC 198  
RESULT 22  
US-09-758-759-128/c  
Sequence 128, Application US/09758759  
Publication No. US20040101832A1  
GENERAL INFORMATION:  
APPLICANT: Hosted, Thomas J.  
APPLICANT: Wang, Tim X.  
APPLICANT: Horan, Ann C.  
TITLE OF INVENTION: Everniminocin Biosynthetic Genes  
FILE REFERENCE: ID0983K US  
CURRENT APPLICATION NUMBER: US/09/758,759  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: US 60/175,751  
PRIOR FILING DATE: 2000-01-12  
NUMBER OF SEQ ID NOS: 204  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 128  
LENGTH: 858  
TYPE: DNA  
ORGANISM: Micromonospora carbonacea  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(858)  
OTHER INFORMATION: evbt  
US-09-758-759-128

Query Match 9.4%; Score 49; DB 12; Length 858;  
Best Local Similarity 51.1%; Pred. No. 0.0045;  
Matches 115; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
QY 29 CGGGCGGCCATCAGAACTGCCTCGCGGGGTGACAGGTGACACAGGTGGAGTT 88  
DB 720 CCCCTGAGTAGACGGTGGCGGCTCGTGGCTGCGGAGCGCGGAGGAGTC 661  
QY 89 GAGCAGCAGCTCATTCGCGGCTGCGCGGATGCGGAAACACCGGGCCAGGATCAGTG 148  
DB 660 GAGCAGCAGAACTGCGCGGCTCGCGGCGCCAGCGGCGCGCGCGGCGAGCAC 601  
QY 149 CAGCAGCGGATGCGCGCTCGATCTCGACGCTGCGCGGAGCTGATCTCGTCCCGG 208  
DB 600 CCGTACGGTATGTTGTCGGCTTCCCGCAGCAGGAGTCCCGAGCTGTCGGCCAT 541  
QY 209 CTCCACACACAGGGAAACTGCTCGGTGGCAGCGGCGAGCCCCAG 253  
DB 540 CACCGATCGGCCCGGACCGGGCGGCGAGCACCGCCTCGCTGAG 496

RESULT 23  
US-10-152-886-14/c  
Sequence 14, Application US/10152886  
Publication No. US20030064491A1  
GENERAL INFORMATION:  
APPLICANT: ECOPIA BIOSCIENCES INC.  
APPLICANT: Farnet, Chris  
APPLICANT: Staiffa, Alfredo  
APPLICANT: Zazopoulos, Emmanuel  
TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDIYNE RING  
FILE REFERENCE: 3011-3US  
CURRENT APPLICATION NUMBER: US/10/152,886  
CURRENT FILING DATE: 2002-05-21  
NUMBER OF SEQ ID NOS: 102  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 14  
LENGTH: 5760

```
; TYPE: DNA
; ORGANISM: Micromonospora echinospora calichensis
US-10-152-886-14

Query Match          9.4%; Score 49; DB 13; Length 5760;
Best Local Similarity 45.0%; Pred. No. 0.0034;
Matches 232; Conservative 0; Mismatches 275; Indels 8; Gaps 1;

QY 10 CGAAGACCGGTGGTGGCCCGCGCGCCCATCAGAACTCCGCGGGGGTGACCA 69
Db 5440 CGAGCGTCCCGTCCCGCGCGAGCGGAGGTCCAGCGCGCCGAGTGCACCACCGAGACGG 5381
QY 70 GGTCCAGACAGGTGGAGTTGACGAGCCAGCTCATCGCGCCCTGCGCCCGCATGCCGAAC 129
Db 5380 TGTAAGGAGTTCGACCTCCGCGCGCCGCTCGGGCGGCTGCGCACACCGAGCTGGC 5321
QY 130 ACCGGCCAGATACAGTTCAGACAGCGCGATCGCGCTCGATCTCGACGGTTCGCGCGCA 189
Db 5320 CCAGCGCGGGAGAGCAGCGCGCGGAGTTCGCGGGGCGCGCCCGCGGTGCGGGCGGA 5261
QY 190 GTCGATCTCGTCCCGCGCTCCACACACGAGGGGAACTGGCTCGGTGGCAGCGGACGCC 249
Db 5260 CCGGATGTTCGCGGGGACACTCGGCGAGCGCTGTTGACGAGCGGCGCCCGACAGCG 5201
QY 250 CCAGCGCGGACAGCTGCTCCAGCTGGCTACGAGCTCTTCGCGGGTTCGGAGCGGAAA 309
Db 5200 CGCGCGGCGAGACCGCGCGCGGTGCGGAGCGACGCGCGCGGCGAGACCCATCC 5141
QY 310 CGCGCAGCGCTCAGATCCCTGTCAGTTCGATCGGCTCAGTGCCTGCTGCTCCCTTGGCC 369
Db 5140 AGTCTCGATACCGGCGCGCGTGGCTGCGGAGCAGACGCTCGTACAGGTAGGCGGTGC 5081
QY 370 TGGGAGGATAGCGGTTTACGACGAGCGGCACACGCGCGCGCGGGGCGC-----GGGGG 421
Db 5080 CGTCTGTTTCCGCTTCCCGCGCGGTCTATGTCACGCGCTCGGTGCGCGAGCTTGGGGC 5021
QY 422 CGGTTACGCGGATCCCTCGATGATACGAGCGGTTCGGGGTTCGGGGCGGTTCGCGAGATCC 481
Db 5020 CGCGCAGCAGATTCGCTCGATGCTCGGCTCGCGGAGCAGCGCTCGGCTCGCGGACGACCT 4961
QY 482 GTACCGCGCGGACCGCGCTCGGCGAGCGCGCGCGCGGG 516
Db 4960 GGATCCCGTGCATGACGCGTTCGCGGCGCGCGGG 4926

RESULT 24
US-10-132-134-13/c
; Sequence 13, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 24081
; TYPE: DNA
; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-13

Query Match          9.4%; Score 49; DB 15; Length 24081;
Best Local Similarity 49.8%; Pred. No. 0.0027;
Matches 124; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 71 GTCCAGCAGGTGGAGTTGAGCAGCCAGCTCATCGCGCTCGATCTCGACGCTCGCGCGCAG 130
Db 30119 GTGCGCTGTGGAGTTGCGATGAGCTCACGCGAGGAGATCAGCAGGACGAAGTC 30060
QY 131 CCGGCGCCAGGATCACGTGCGAGCAGCGGATGCGCGCTCGATCTCGACGCTCGCGCGCAG 190
Db 30059 CAGCCCGTCCCGCGGAGACGTCGCGGAGTGGCGCAGCGGCTCGACCTTGGCGCGAG 30000
QY 191 CTGATCTGTGCGCCCGGCTCCACACCGAGGGAACTGGCTCGGTGGGAGCGGCGAGCCC 250
Db 29999 GGTGGCTTCCAGTTTCGGCTCGGTTCATCCGCGGCGAGCGCGGTGTGACTGAACACCATCGC 29940
QY 251 CAGCGCGGACAGCTCGTCCAGGCTGGGTTCACGAGCTCTCGCGGGTTCGGGAGCGGAAAC 310
Db 29939 CAGGTGCACACCGCGCTCGATGGCGGCTAGCGGCGGAGCAGACCTCGTCTCCGCTCGCGGC 29880
QY 311 GCGCAGCGC 319
Db 29879 GAGCGCGTC 29871

RESULT 26
US-09-758-759-1/c
; Sequence 1, Application US/09758759
; Publication No. US20040101832A1
; GENERAL INFORMATION:
; APPLICANT: Hosted, Thomas J.
; APPLICANT: Wang, Tim X.
; APPLICANT: Moran, Ann C.
; TITLE OF INVENTION: Everninomycin Biosynthetic Genes
```

FILE REFERENCE: ID0983K US  
CURRENT APPLICATION NUMBER: US/09/758,759  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: US 60/175,751  
FILING DATE: 2000-01-12  
NUMBER OF SEQ ID NOS: 204  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 1  
LENGTH: 109519  
TYPE: DNA  
ORGANISM: Micromonospora carbonacea  
US-09-758-759-1

Query Match 9.4%; Score 49; DB 12; Length 109519;  
Best Local Similarity 51.1%; Pred. No. 0.0022;  
Matches 115; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 29 CCGGCGCCCATCAGAACTCCACTCGCGCGGGTGACACAGTGCACCGAGTGGAGTT 88  
DB 82792 CCCCTGACGTAGACGGTGGCGGCTCGTGGCTGCGGAGCCGAGAGCGGGAAGTC 82733

QY 89 GAGCAGCCAGTCTATCGCGGCTGCGCGGCATGCCGAAACACCGCGCCAGGATCACGTG 148  
DB 82732 GAGCAGCAGAACTGCGCGGCTGCGCGGCAGCGCGCCGCGGCGCAGCGCAC 82673

QY 149 CAGCAGCGGATCGCGGCTCGATCTCAGCGTGGCGCGGAGCTCGTCTGCGCGG 208  
DB 82672 CCGTACGGTGTGTGTCGGCTTCCGACGCGCAGGAGTGGCGGAGCTGTGCGCAT 82613

QY 209 CTCCACACACAGCGGAACTGCTCGTGGCAGCGCGGAGCCCGCAG 253  
DB 82612 CACCGATCGCGCGGCGCAGCGCGCGCGCAGCAGCGCTCGCTGAG 82568

RESULT 27  
US-10-156-761-6291/c  
Sequence 6291, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 6291  
LENGTH: 1116  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1) (1116)  
US-10-156-761-6291

Query Match 9.4%; Score 48.8; DB 15; Length 1116;  
Best Local Similarity 49.4%; Pred. No. 0.0048;  
Matches 159; Conservative 0; Mismatches 157; Indels 6; Gaps 1;

QY 39 ATCAGAACTGCACCTCCGCGGGTGACACAGTGCACCGAGTGGAGTTGACGACGAC 98  
DB 524 ACCATCGGTAGCGTTCGAGCGGAAGATCACGTGACCTGATCCGAGGCGG 465  
QY 99 CTCATCGCGCGCTGCGCGGCATGCCGAAACACCGCGGCGCAGGATCACGTGACGACGCG 158

DB 464 ATGACCTCGTGATCGCGGGATGCCCGCTTGGTGGCGGGATCTTGATGAGCGTGTG 405  
QY 159 ATGCGCGCTCGATCTCGACGCTCGCGCGCAGCTCGATCTCGTCCGCGGCTCCACACC 218  
DB 404 GGGCGGTGACGACGACG-----GCCAGCTCTTGGGCTCGGCGACCGTCCCTTGGTG 351  
QY 219 AGGGAAACTGGCTGGTGGCAGCGGACGCCCGCGGACAGCTCGTCCAGGCTGGCG 278  
DB 350 TTGTGGCCAGCGCGGTGCGACCTCGATCGAGACCCCGCGCTCTGCGCGTGGCG 291  
QY 279 TCAGGACCTCTCGCGGTGCGGAGGGAAGGCGGACGCTCAGATCCCTGTCAGTGGC 338  
DB 290 TCGAAGACCGCGCGGAGGATGTCGGCGGCGTCCGCGAGCTCCGCGTGTGATCATGGG 231  
QY 339 ATCGCTCAGTCCCGTCTGCTCC 360  
DB 230 ATGGCTTCTTCGACGGTGACCC 209

RESULT 28  
US-10-282-122A-14525  
Sequence 14525, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patent In version 3.1  
SEQ ID NO 14525  
LENGTH: 2208  
TYPE: DNA  
ORGANISM: Burkholderia mallei  
US-10-282-122A-14525

Query Match 9.4%; Score 48.8; DB 13; Length 2208;  
Best Local Similarity 49.6%; Pred. No. 0.0044;  
Matches 125; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 65 GACAGGTGACACAGGTTGGAGTTAGACACCGAGCTCATCGCGCTCGGCGGATGCC 124  
Db 826 GACTGGGGGACACGTCGAGCAACTTCTGCGCTCGGCGGACCGTGTCCGATCCG 885  
QY 125 GAAACACCGGCGAGGATCAGTCAGCAGCGGATGCGGCGCTCGATCTCGAGCGTCCG 184  
Db 886 AAGCGACGATTCGGAGCAGCTGTACAGCGCGGATGCGAATCTCGACCATTAACCGCAAG 945  
QY 185 CCGCAGCTCATCTCGTCGCGCGCTCCACACAGGGGAACTGGCTCGGTGGCAGCGG 244  
Db 946 AAGCAGTGTGCTCGGCTATCAGTTCGAGCAACAAGTGAACCGGTGTGGAGCTGCGG 1005  
QY 245 CAGCCCGACCGGACAGCTGTCAGCTGAGCTGCGGTACAGCACTCTCGCGGGTGGAGAC 304  
Db 1006 CAGAAGCTGGCTGGATGCACTGTGCTCGACGAGCGGTGCTCTACGCGCGCGCTC 1065  
QY 305 GGAAAGCGGCAC 316  
Db 1066 GACGACGACAC 1077

## RESULT 29

US-10-156-761-7305/c  
; Sequence 7305, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7305  
; LENGTH: 654  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(654)  
US-10-156-761-7305

Query Match 9.3%; Score 48.6; DB 15; Length 654;  
Best Local Similarity 48.7%; Pred. No. 0.0058;  
Matches 132; Conservative 0; Mismatches 139; Indels 0; Gaps 0;  
QY 17 CCGGTGGTCCCGCGCGCGCCATCAGAACTGCCACTCCGCGGGGTGACCAGTCCGAC 76  
Db 642 CCCTCGCGCGAATCGTTCTCTGGCGCCAGGCGCGCTAGCGCCAGCAGCGCGGTCCAG 583  
QY 77 CAGGTGGGAGTTGAGCAGCAGCTCATCGCGCGCTCGCGCGGCATGCGGAAACACCGGGC 136  
Db 582 CAGGACCGTGGAGAGCAGCGCGGTTCGGCCACCTCCAAACGGGTCTCGATCCGCTCCAG 523  
QY 137 CAGGATCAGTGAGCAGCGCGATGCGCGCTCGATCTCGACGGTTCGCGCGGAGCTCGAT 196  
Db 522 TTGGTCCAGGTCCAGGACCGCGCTCGTTCCGCGAGATCGCGGTACGACAGCAGAGCGCG 463  
QY 197 CTCGTCCCGCGGTCCCAACACAGGGAACCTGGTTCGGTGGCAGCGCGACGCCCGCGCG 256  
Db 462 CCGGTCCAGCGGGCGCGCAGCGTTCGCGACGAGGTTCAGGGTCTCTGCGAGCGCGGACCT 403  
QY 257 GGACAGCTCGTCAGGCTGGCGTTCACGGACC 287

Db 402 GCGCGCGCGCGACTTGTGGACGTCCAGCCC 372

## RESULT 30

US-10-282-122A-28770/c  
; Sequence 28770, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 28770  
; LENGTH: 1797  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-10-282-122A-28770

Query Match 9.3%; Score 48.6; DB 13; Length 1797;  
Best Local Similarity 47.7%; Pred. No. 0.005;  
Matches 173; Conservative 0; Mismatches 189; Indels 1; Gaps 1;  
QY 154 CCGCGATCGCGCTCGATCTCGACGCTCGCGCGCAGTCGATCTCGTCGCGCGCTCCC 213  
Db 1768 CGCGCTGCTCGCCCCGATCAACAGATTCTTCTGGGAATACCGGCGCCCGGTGATCG 1709  
QY 214 ACACAGGGAACTGCTCGGTGGCAGCGGAGCCCGACCCCGGAGACAGCTCGTCCAGGC 273  
Db 1708 GCAACGGCACAACCGGCGCCCCCGGACACCGGGCCAAATGGCGGGACGGCGGTGTTAA 1649  
QY 274 TGGCGTCACGAGCTCTCGCGGGTTCGGAGCGGAAACGCGCACGCGTCAAGTCCCTGTCA 333  
Db 1648 TCGCAACCGGCGCGCGGAGGATTCGCGCGCGCGCGCGGTCAACGCGGCGCGCGGCGCA 1589  
QY 334 GTGCAATCGGCTCAGTTCGCGGTCTGTCCTTGGGAGGATAGCGGTTCACGACGA 393  
Db 1588 ACGCGCGCGCGCGGCTCATCGGCAACGCGCGCGCGCGCGCGCGCGCGGAGGCGCA 1529



QY 123 CGAAACACCGGCGCAGGATCAGTCAGCAGCGGCGATCGCGCCTCGATCTCGACGGTC 182  
Db 3479 CTGGCGGCGGATACATCCGCTTCAGCGCGACCTGGGCGGATCGGCGAGGTCGATCGGC 3420  
QY 183 GCGCGAGCTCGATCTCGTCGCGCGGCTCCCHACACAGGGGAACTGGCTCGTGGGAGC 242  
Db 3419 GAGAGCGGAGCGGTGGCGCGCGACCGCGCGCGACACCTCGGTGATCGCGCGCGCGCAC 3360  
QY 243 GCGAGCGCGCGGCGGACAGCTCGCTCCAGGCTGGGCTCAGGACCTCTCGCGGGTCGGGA 302  
Db 3359 AGCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3300  
QY 303 GCGGAAACCGCGCAGGCTCAGATCCCTGTGATCGCATCGGCTCAGTCCGGTCGTGCC 362  
Db 3299 CGGTGACGAGTCTCTCGACCAACCGC-GTCCGATCTCGTCTCGGAGCGCGCGCGCGCAG 3241  
QY 363 C-----TTGGGCTGGGAGATAGCGGTTTCAGACGAGCGGCGGCGGCGGCGGCGG 416  
Db 3240 CGCGGATAGCGCTCCACCGCGAGCGGATCGTGGCGATGATGTCGCGGCGGAGGTGCGC 3181  
QY 417 GGGGCGGCTTCAGCGGATCCGTCGATGACACGCGCTCGCGGGTCGCGGGCGGTCGGCGA 476  
Db 3180 GTCCACCGCGAGGTGATCCGCGCGCTCTTGGGCTCTCTCGCGGCTGATGAGGCTCGGCC 3121  
QY 477 GATCGTACCGCGCGGCGCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 520  
Db 3120 GTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3077

## RESULT 33

US-10-369-493-41689/c  
; Sequence 41689, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 41689  
; LENGTH: 1350  
; TYPE: DNA  
; ORGANISM: SPHINGOMONAS  
US-10-369-493-41689

Query Match 9.3%; Score 48.2; DB 16; Length 1350;  
Best Local Similarity 47.2%; Pred. No. 0.0066;  
Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;  
QY 64 TGACGAGTTCGACGAGTGGGAGTTGAGCAGCAGCTCATCGCGGCTCGCGCGCGCATGC 123  
Db 1108 TGGCAGCTCGGCGATCTGTCGAGCGCGGAGAACCCCGCCACGCGCTCGGACGACAGTC 1049  
QY 124 CGAAGACCGGCGGAGGATCAGTCGAGCAGCGGATCGGCGCTCGATCTCGAGGTCG 183  
Db 1048 CGGTATGCGGCTCAGGATCCGCGCGCGCGGCTCGGCGGCGGAGACCTTGGCGAGCGCGC 989  
QY 184 GCGCAGCTCGATCTCGTCCGCGGCTCCACACGAGGAGAACTGGCTCGGTGCGCAGCG 243  
Db 988 GGGGAGCAAGCGGTCAGCGCGGATCAGAGAGATCGGCTCTGGCGGTACATCATGA 929  
QY 244 GAGCGCGCGCGGAGACGCTCGTCAGCTGGGCTCAGGACCTCTCGCGGTCGGAG 303  
Db 928 CCAGGATCACGAGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGG 869

QY 304 CGAAACCGCGCAGGCTCAGATCCCTGTCAGTCGATCGGCTCAGTCGCGGTCGTCGCCCC 363  
Db 868 GTGGCGGCGGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 809  
QY 364 TTGGCCTGG 372  
Db 808 TCGGCGATCG 800  
RESULT 34  
US-10-338-110-119  
; Sequence 119, Application US/10338110  
; Publication No. US20040023254A1  
; GENERAL INFORMATION:  
; APPLICANT: Fuhrmann, Jeffrey J.  
; APPLICANT: Romesser, James A.  
; TITLE OF INVENTION: A Method To Assess Quorum Sensing Potential Of Microbial  
; FILE REFERENCE: HER-0056  
; CURRENT APPLICATION NUMBER: US/10/338,110  
; CURRENT FILING DATE: 2003-01-07  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 119  
; LENGTH: 536  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus Sequence  
; NAME/KEY: misc feature  
; LOCATION: (7)..(7)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (21)..(21)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (33)..(33)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (69)..(69)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (87)..(87)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (213)..(213)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (226)..(226)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (261)..(261)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (273)..(273)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (347)..(347)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (357)..(357)



```

; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (398)..(399)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (419)..(419)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (509)..(509)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (517)..(517)
; OTHER INFORMATION: n is a, c, g, or t
US-10-338-110-119

Query Match
Best Local Similarity 9.2%; Score 48; DB 17; Length 536;
Matches 59; Conservative 185; Mismatches 170; Indels 0; Gaps 0;

QY 68 CAGGTCGACGAGTGGGAGTTGACGACGACGATCATCGCGCTCATCGCGCGGCGATCGCGAA 127
DB 1 SVRTKVMVMGVTWTERHNSYRHKCTSGSNCRCTWYCGGCRYSRVSVWTTTYSRR 60

QY 128 ACACGGGCCAGGATCAGTGCAGACGCGGATCGCGCTCGATCTCGACGTCGCGCG 187
DB 61 MRRCTVGNTGGSVRSYBYCVWYSKMMVVDMMBSKMWGARYKBGAYCARTWCGAYMVY 120

QY 188 CAGCTCATCTCGTCGCGCGGCTCCACACAGGGGAACTGCTCGTGGCAGCGGCG 247
DB 121 SMSAHRKSSKCTACVTSTBSGSHKRVGMVSVTVGBDHATSKYGGHTGYGSCGSGTG 180

QY 248 CCCAGCGCGGACAGCTCGTCAGGCTGGCGTCACGACCTCTCGCGGTCGCGAGCGGA 307
DB 181 HTDCBACACVBDSSCBWACTCTBRARKVNSTVTYRSBMYNSYBKCBSVSRVM 240

QY 308 AACGGCAGCGCTCAGATCCTCTCAGTCGATCGGCTCAGTCGCGTCTGCTCCCTTGG 367
DB 241 MYRCSSMVATCCVRSRNTTGGARHTWCNCGMTWCRCRMCRCRHSKSSAMYKGG 300

QY 368 CCGGAGGATAGCGGTTTCAGACGAGCGGACACACGCGCGGCGGCGGCGGCGGCTTC 427
DB 301 CSRTGMVKWRTYTSKSDVMBBCSTSVARVMGSCDSSVYVSVBSGGNGCVVRKSVNTR 360

QY 428 AGCGATCGCTCATACACGAGCGCTGCGGGTTCGCGGGGTCGCGGATCC 481
DB 361 TSGSVRSYRSHCRBSRSTAGWCVGSYKTKWCDVMKMSAMGGBRTYMSB 414

RESULT 35
US-10-200-562-193/c
; Sequence 193, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowan, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200.562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 3957
; TYPE: DNA
; ORGANISM: HSV2
US-10-200-562-193

Query Match
9.2%; Score 48; DB 15; Length 3957;
Best Local Similarity 46.6%; Pred. No. 0.0063;
Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;

QY 23 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82
DB 3033 GTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2974

QY 83 GGAGTTGAGCAGCAGCAGCTCATCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142
DB 2973 GCAGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2914

QY 143 CACGTGACAGCAGCGGATGCGCGCTCGATCTCGACGTCGCGCGCGCGCGCGCGCGCG 202
DB 2913 CGCGCGCAGGTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2860

QY 203 GCCCGGCTCCACACACGAGGGAATGCTCGGTGCGAGCGCGCGCGCGCGCGCGCGCG 262
DB 2859 CTCGCGGTGCGGCACTGCGCATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2800

QY 263 CTCGTCCAGGTGCGGTCAACGACCTCTCGCGGTCGCGGTCGCGGAAACGCGCGCTCA 322
DB 2799 CAGCGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2740

QY 323 GATCCTCTGTCAGTCGCTCAGTGCCTGCTCGTCCCTTGCCTTGCCTTGGGAGGATACG 382
DB 2739 CGAGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2680

QY 383 GTTCACGACGAGCGGACACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442
DB 2679 GTGGTCGTGAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2620

QY 443 TGACCA-----GCGGCTGCGGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496
DB 2619 GGGCGCGCGGTGTGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2560

QY 497 CCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516
DB 2559 GCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2540

RESULT 36
US-10-237-551-193/c
; Sequence 193, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237.551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 3957
; TYPE: DNA
; ORGANISM: HSV2
US-10-237-551-193

Query Match
9.2%; Score 48; DB 15; Length 3957;
Best Local Similarity 46.6%; Pred. No. 0.0063;
Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;

QY 23 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82
DB 3033 GTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2974

QY 83 GGAGTTGAGCAGCAGCAGCTCATCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142
DB 2973 GCAGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2914

Query Match
9.2%; Score 48; DB 15; Length 3957;
Best Local Similarity 46.6%; Pred. No. 0.0063;
Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;

QY 23 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82
DB 3033 GTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2974

QY 83 GGAGTTGAGCAGCAGCAGCTCATCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142
DB 2973 GCAGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2914
```

```
QY 143 CAGCTCAGACGCGGATCGGCGCTCGATCTCGACGCTCGGCGCGGATCGATCTCGTC 202
Db 2913 CCGCGCCAGGTCCTCCCGCGGAGCGCGAGTAGAGAT-----CACACCGCACGTC 2860
QY 203 GCCCGGCTCCACACACAGGAGAACTGGCTCGGTGCGACGCGGAGCCCGGAGCGGACAG 262
Db 2859 CTCGGGTCGGGACCTTCGCGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2800
QY 263 CTGCTCAGAGCTGGCTCAGGACCTTCCTCGGCTCGGAGCGGAGAAACCGCGACGCTCA 322
Db 2799 CAGCGGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2740
QY 323 GATCCCTGTGTCAGTCGATCGGCTCAGTCGCGGTTCGCTCCCTTCGCTGCGAGTAGCG 382
Db 2739 CGAGGCGAGCGCGCGGTCGACATAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2680
QY 383 GTTCACGACAGCGGACCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442
Db 2679 GTGTCCTGTGAGCTCGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2620
QY 443 TGACCA-----CGCGCTCGGCGGTGCGGCGCGGTGCGGCGCGGTGCGGCGCGGT 496
Db 2619 GGGCGCGGCGGTGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2560
QY 497 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516
Db 2559 GCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2540

RESULT 37
US-09-827-688-8
; Sequence 8, Application US/09827688
; Publication No. US20030165476A1
; GENERAL INFORMATION:
; APPLICANT: ORSON, FRANK
; APPLICANT: KINSEY, BERMA
; APPLICANT: BHOGAL, BALBIR
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION
; FILE REFERENCE: P01949US1/10004014
; CURRENT APPLICATION NUMBER: US/09/827,688
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,680
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Query Match 9.2%; Score 48; DB 10; Length 154746;
Best Local Similarity 46.6%; Pred. No. 0.0037;
Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;

QY 23 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82
Db 129000 GTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129059
QY 83 GGAATTGAGCAGCAGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142
Db 129060 GCAGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129119
QY 143 CAGCTCAGACGCGGATCGGCGCTCGATCTCGACGCTCGGCGCGCGCGCGCGCGCGCG 202
Db 129120 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129173
QY 203 GCCCGGCTCCACACACAGGAGAACTGGCTCGGTGCGACGCGGAGCCCGGAGCGGACAG 262
Db 129174 CTCGGGTCGGGACCTTCGCGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129233
QY 263 CTGCTCAGAGCTGGCTCAGGACCTTCCTCGGCTCGGAGCGGAGAAACCGCGACGCTCA 322
```

```
Db 129234 CAGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129293
QY 323 GATCCCTGTGTCAGTCGATCGGCTCAGTCGCGGTTCGCTCCCTTCGCTGCGAGTAGCG 382
Db 129294 CGAGGCGAGCGCGCGCGCGGTTCGACATGAGGCGCGCGCGCGCGCGCGCGCGCG 129353
QY 383 GTTCACGACAGCGGACCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442
Db 129354 GTGTCCTGTGAGCTCGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129413
QY 443 TGACCA-----CGGCTGCGGCGGTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGT 496
Db 129414 GGGCGCGCGGTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129473
QY 497 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516
Db 129474 GCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129493

RESULT 38
US-09-827-688-8/c
; Sequence 8, Application US/09827688
; Publication No. US20030165476A1
; GENERAL INFORMATION:
; APPLICANT: ORSON, FRANK
; APPLICANT: KINSEY, BERMA
; APPLICANT: BHOGAL, BALBIR
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION
; FILE REFERENCE: P01949US1/10004014
; CURRENT APPLICATION NUMBER: US/09/827,688
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,680
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Query Match 9.2%; Score 48; DB 10; Length 154746;
Best Local Similarity 46.6%; Pred. No. 0.0037;
Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;

QY 23 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82
Db 152742 GTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152683
QY 83 GGAATTGAGCAGCAGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142
Db 152682 GCAGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152623
QY 143 CAGCTCAGACGCGGATCGGCGCTCGATCTCGACGCTCGGCGCGCGCGCGCGCGCGCG 202
Db 152622 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152569
QY 203 GCCCGGCTCCACACACAGGAGAACTGGCTCGGTGCGACGCGGAGCCCGGAGCGGACAG 262
Db 152568 CTCGGGTCGGGACCTTCGCGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152509
QY 263 CTGCTCAGAGCTGGCTCAGGACCTTCGCGCGCGGTTCGCGCGCGGTTCGCGCGCGGT 322
Db 152508 CAGCGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152449
QY 323 GATCCCTGTGTCAGTCGATCGGCTCAGTCGCGGTTCGCTCCCTTCGCTGCGAGTAGCG 382
Db 152448 CGAGGCGAGCGCGCGCGCGGTTCGACATGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 152389
QY 383 GTTCACGACAGCGGACCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442
```

Db 152388 GTGTCGTGAGCTCGGCACCGCCCGGGGGCGCAGTAGGCTCCAGGGCGCGCGCGGA 152329  
Qy 443 TGACCA-----GGGGTTCGGGGTTCGGGGGGTTCGGGGAGATCCGTACCGCCCGACCG 496  
Db 152328 GGGGGCGGGGTGTGGTGGGGCCCGCGGTGGCGCGCGCCAGCGCGCTTCGGGGTCGGG 152269  
Qy 497 CCTGGCCAGCGCGCGGG 516  
Db 152268 GCGCTCGGGCGCGCGGG 152249  
RESULT 39  
US-10-156-761-2330/c  
; Sequence 2330, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2330  
; LENGTH: 1341  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1341)  
US-10-156-761-2330  
Query Match 9.2%; Score 47.8; DB 15; Length 1341;  
Best Local Similarity 46.0%; Pred. No. 0.0083;  
Matches 199; Conservative 0; Mismatches 232; Indels 2; Gaps 1;  
Qy 76 CAGGTGGAGTTGAGCAGCGAGTCAATCGCCGCTCGCGGATCGCGGATCCGAAACACCGGG 135  
Db 766 CCGAGCGGGCCACCGGCAAGTCCGAGCGGTTCAGGAAACCGAGGTTGACCGCGCGCGCGC 707  
Qy 136 CCAGATCAGTGCAGCAGCGGATGCGCGCTCGATCTCGACGTCGCGCGCGCAGCTCGA 195  
Db 706 CGATCCCGGGGAGGAGCGTGTGGCGGCGGACGCGGCGAATGACACCGCACCGACC 647  
Qy 196 TCTCGTCCCGGCTCCACACACAGGGGAAACTGGCTCGTGGCAGCGGCGAGCCCGACGC 255  
Db 646 GCGCGCGGGCAGCGGTCGCGCACCGCACCTGTGGCGAGCGCGCACACACCCACG 587  
Qy 256 GGGACAGTCTCCAGGTGGCGTCAAGGCTCTCGCGGTTCGGGAGCGGAGCGGAAACGCGCA 315  
Db 586 TGACACCGCGCGCTCGCGGCCACACAGACGACCGGTCGCGGCGCGGCGAGCGCGT 527  
Qy 316 CGGCTCAGATCCCTCTCAGTCGCTCGGCTCAGTCCGCTCGTCCCTTGGCCCTGGGAG 375  
Db 526 CGCGCGCGGCTCGATGACGCGGCCAGTGTTCATGGCGA--GAAAGACGGCGCGACGAG 469  
Qy 376 GATAGCGTTTACGACGAGCGCACCGCGGGGGGGGGGGGGGGGGGGTTCAGCCGATC 435  
Db 468 GACGAGCAGCGCGGAGCGGCTGTCACGCGGAGCGGGTCCGTGCGGCGCTCGGCGCGCTC 409  
Qy 436 CGCTCGATGACACGAGCGGTTCGGGGTTCGGGGCGGTTCGGCGAGATCCGTACCGCCCGGACC 495  
Db 408 CAGCGCTCGGGGGGTTCGGGGCGCTCGGGCGCTCGGGCGCGTTCGGGGCGCGCGGGGC 349

Qy 496 GCCTCGGCACGG 508  
Db 348 GCGCGGGCGCGG 336  
RESULT 40  
US-10-156-761-7034/c  
; Sequence 7034, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7034  
; LENGTH: 2997  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2997)  
US-10-156-761-7034  
Query Match 9.2%; Score 47.8; DB 15; Length 2997;  
Best Local Similarity 45.2%; Pred. No. 0.0074;  
Matches 214; Conservative 0; Mismatches 257; Indels 2; Gaps 1;  
Qy 1 GGTGACGACGAAACGACCGGTGGTTCGCGCGCGCGCGCGCATCAGCACTGCCACTTCGCGG 60  
Db 1780 GCTCCAGCTCCGCTCCCTCGCGAGTGGCGGTGGCGCGGTTCGCGCGCGGTTCGCGG 1721  
Qy 61 GGTGACGACGTCGACGAGTGGAGTTGAGCAGCGAGTTCATCGCCGCTTCGCGCGGCA 120  
Db 1720 TCGCGCGCGCGAGTTCCTCGCGTACGAGCGCGAGTTCGCGCTTCGCGCGGTTCGCGG 1661  
Qy 121 TGCGGAAACACCGCGCGCGAGTACGTGACGACGCGCGATGCGCGCTTCGATTCGACGG 180  
Db 1660 CGTCGCGCGGTGGTAGGCGGTGAGGCGGTGAGGCGCGCTCTCGGTCTCCGCTCGACATGCCGG 1601  
Qy 181 TCGCGCGAGTTCGATCTC--GTGCGCGGTTCGCCACACGAGGAGAACTGGCTCGGTGG 238  
Db 1600 TGATCTTCGCGCGCGCGCGCGATGTCGAGGAGCGCGATGTCGAGGAGCGCGACACGCGGAGCGCGT 1541  
Qy 239 CAGCGCGAGCGCGCGGACAGTTCGTCAGGCTGGCGTTCAGCGACCTCTTCGCGGGTTC 298  
Db 1540 CGACGAGACCGCGCGGAGTTCGCGCGGATGCCCTGAAGGCACTGTTCTTCAGGTGCA 1481  
Qy 299 GGGAGCGGAAACGCGACGCGCTCAGATCCCTGTGATGCGATCGGCTCAGTCCGCTCGT 358  
Db 1480 GCCAGTGGACCGCGCGGTTCGCTCGCGCGCGCGAGGCGCGAGGCGCGCGCGCGT 1421  
Qy 359 CCGCTTCGCGTGGAGGATAGCGGTTTACGACGAGCGCGCACCGCGCGCGCGCGGCGG 418  
Db 1420 CGTCCGTGCTTCGCGGAGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1361  
Qy 419 GGGCGGTTTTCAGCGGATTCGCTCGATGACAGCGGCTTCGCGGCTTCGCGGCGGTTC 471  
Db 1360 GGTTCGCGCGCGCGCGAGTCTCGCGCGCGGTTCGCGCGCGCTTCCTTCGCGCGAC 1308

RESULT 41

US-10-156-761-3185/c  
; Sequence 3185, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 3185  
; LENGTH: 3789  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS (3789)  
; LOCATION: (1)..(3789)  
US-10-156-761-3185

Query Match 9.2%; Score 47.6; DB 15; Length 3789;  
Best Local Similarity 46.5%; Pred. No. 0.008;  
Matches 188; Conservative 0; Mismatches 214; Indels 2; Gaps 1;  
QY 115 CCGGATCCGGAACACCGGCGCAGATCAGTGCAGCGGATGGCGCTCGATCT 174  
DB 901 CCGGATCCGGAACACCGGCGCAGATCAGTGCAGCGGATGGCGCTCGATCT 174  
QY 175 CCACGCTCGCGCGCAGTCTCGTCCGCGGCTCCACACGAGGGAACCTGGCTCG 234  
DB 841 CCGCGCGGAGGAGCGGCGCGCGTTCGCGCGGCGGCGCGCGCGCGCGG 782  
QY 235 GTGCGAGCGGAGCGGCGGCGGAGCGTGTTCAGCGTGTTCAGCGGCGGCGG 294  
DB 781 CCGCGAGCGGAGCGGCGGCGGAGCGTGTTCAGCGTGTTCAGCGGCGGCGG 722  
QY 295 GGTGCGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 354  
DB 721 CCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 662  
QY 355 TCGTCCCGCTTGGCTGGAGGATAGCGGTTCCAGCGGCGGCGGCGGCGGCGG 414  
DB 661 TGTCCCGCGGCTAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 602  
QY 415 GCGGCGGCGGCTTCCAGCGGAT--CCGCTCGATGACGAGCGGCTGCGGCGG 472  
DB 601 CCGGCGGCGGCTAGCGGCTTCCAGCGGCGGCGGCGGCGGCGGCGGCGG 542  
QY 473 GCGAGATCGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 516  
DB 541 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 498

## RESULT 42

US-09-758-059-1/c  
; Sequence 1, Application US/09976059  
; Patent No. US2002016474A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Zazopoulos, Emmanuel  
; APPLICANT: Stafra, Alfredo  
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
; FILE REFERENCE: 3019-PCT  
; CURRENT APPLICATION NUMBER: US/09/976,059

CURRENT FILING DATE: 2001-10-15  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 89421  
TYPE: DNA  
ORGANISM: Actinoplanes sp.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (2077)..(3078)  
OTHER INFORMATION: ORF 1; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (3118)..(4032)  
OTHER INFORMATION: ORF 2; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (6665)..(5814)  
OTHER INFORMATION: ORF 4; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (7703)..(6693)  
OTHER INFORMATION: ORF 5; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (9464)..(8130)  
OTHER INFORMATION: ORF 6; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (9691)..(10761)  
OTHER INFORMATION: ORF 7; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (12751)..(10829)  
OTHER INFORMATION: ORF 8; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (13617)..(12802)  
OTHER INFORMATION: ORF 9; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (15203)..(13614)  
OTHER INFORMATION: ORF 10; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (15591)..(15863)  
OTHER INFORMATION: ORF 11; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (15880)..(19035)  
OTHER INFORMATION: ORF 12; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (19032)..(39713)  
OTHER INFORMATION: ORF 13; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (39713)..(65800)  
OTHER INFORMATION: ORF 14; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (65826)..(66530)  
OTHER INFORMATION: ORF 15; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (66546)..(67370)  
OTHER INFORMATION: ORF 16; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (67384)..(70059)  
OTHER INFORMATION: ORF 17; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (70099)..(70662)  
OTHER INFORMATION: ORF 18; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (70659)..(71906)  
OTHER INFORMATION: ORF 19; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (73439)..(71964)  
OTHER INFORMATION: ORF 20; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (74216)..(73563)

;	OTHER INFORMATION: ORF 21;	negative strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (75434)..(74213)	
;	OTHER INFORMATION: ORF 22;	negative strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (75535)..(76464)	
;	OTHER INFORMATION: ORF 23;	positive strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (78110)..(76449)	
;	OTHER INFORMATION: ORF 24;	negative strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (79864)..(78107)	
;	OTHER INFORMATION: ORF 25;	negative strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (81624)..(79861)	
;	OTHER INFORMATION: ORF 26;	negative strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (81909)..(81682)	
;	OTHER INFORMATION: ORF 27;	negative strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (82346)..(82062)	
;	OTHER INFORMATION: ORF 28;	negative strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (82587)..(84446)	
;	OTHER INFORMATION: ORF 29;	positive strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (84481)..(85548)	
;	OTHER INFORMATION: ORF 30;	positive strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (85556)..(86845)	
;	OTHER INFORMATION: ORF 31;	positive strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (87372)..(86803)	
;	OTHER INFORMATION: ORF 32;	positive strandedness
;	NAME/KEY: misc.feature	
;	LOCATION: (87494)..(88420)	
;	OTHER INFORMATION: ORF 33;	positive strandedness
;	US-09-976-059-1	

Query Match	9.1%;	Score 47.4;	DB 9;	Length 88421;
Best Local Similarity	49.8%;	Pred. No. 0.0056;		
Matches 120;	Conservative 0;	Mismatches 121;	Indels 0;	Gaps 0;
Qy	43	CGAACTGCCACTCCGGCGGGGTACCAAGTTCGACACAGGTGGAGTTGACAGCCAGCTCA	102	
Db	59630	CGACGCCGTGCCGATCGGGATGTCGGCGCCGGCCGAGCGGACACGACGCGTGCCGA	59571	
Qy	103	TCGCGCGCTCGCGCGGATGCCGAACACCGGGCCAGGATCAGGTGCAGCACCGCGATGC	162	
Db	59570	AGCGCGCTTCGACACCAATGAACACGGTCACGCCCTTCGACCGGGCCAGCTCGGCCAGGC	59511	
Qy	163	GCSCCTCGATCTCGACGGTCGGCGCGCAGCTCGATCTCGTCGCCCGCCGCTCCACACACAGGG	222	
Db	59510	GCCGGTGCACTTCGGCGCGCACCCGACCGGCACCTCGTGTCGCCGTCGGTGCCCTCGG	59451	
Qy	223	GAAACTGGCTCGGTGSCAGCGCAGCCCGACCGGACAGCTCGTCAGGCTGGCGGTAC	282	
Db	59450	CCGGCGCGCGCCGGTCGGCGCGGAGTTCAGTCTCTCCGGTACGCCGTCGACGCGCTCCC	59391	
Qy	283	G 283		
Db	59390	G 59390		

RESULT 43  
US-09-893-519A-87/c  
; Sequence 87, Application US/09893519A  
; Publication No. US20030027243A1  
; GENERAL INFORMATION:  
; APPLICANT: ANADYS PHARMACEUTICALS, INC.  
; APPLICANT: THOMPSON, Craig  
; APPLICANT: MOORE, Jeffrey  
; APPLICANT: BUURMAN, Ed T.

```

? APPLICANT: BRADLEY, John
? APPLICANT: DESILVA, Thamara
? APPLICANT: HARRIS, Sandra
? APPLICANT: KOMARNITSKY, Svetlana
? APPLICANT: MENDILLO, Marc
? APPLICANT: MOORE, Daniel
? APPLICANT: MCCOY, Melissa
? APPLICANT: SANDERSON, Karen
? APPLICANT: HAQ, Tariq
? APPLICANT: ZHU, Shuhao
? APPLICANT: LONG, Fan
? APPLICANT: DAVIDOV, Eugene
? TITLE OF INVENTION: ANTI-FUNGAL COMPOUNDS AND METHODS OF USE
? FILE REFERENCE: 0342/1GS48-US2
? CURRENT APPLICATION NUMBER: US/09/893,519A
? CURRENT FILING DATE: 2001-06-28
? PRIOR APPLICATION NUMBER: US 60/215,164
? PRIOR FILING DATE: 2000-06-29
? PRIOR APPLICATION NUMBER: US 60/224,457
? PRIOR FILING DATE: 2000-08-10
? NUMBER OF SEQ ID NOS: 146
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 87
? LENGTH: 2307
? TYPE: DNA
? ORGANISM: Homo sapiens
? PUBLICATION INFORMATION:
? DATABASE ACCESSION NUMBER: Human Genbank/Y11354.1
? DATABASE ENTRY DATE: 1997-06-25
? RELEVANT RESIDUES: (1)..(2307)
? US-09-893-519A-87

```

Query Match	9.1%	Score 47.2	DB 10	Length 2307
Best Local Similarity	44.6%	Pred. No. 0.011		
Matches 232	Conservative 0	Mismatches 283	Indels 5	Gaps 1
Qy	2	GTGCAGACGACGACCGGTTGTCGCCCGCGCGCCCATCAGCACTGCCACATCCCGCGG 61		
Db	1269	GGGGGTCGCGGACAGGCTCTGGGTCACTGGCGCGCGCGCCCTTTGGCGACCGCGGTGG 1210		
Qy	62	GGTGTACAGGTCCACAGGTGGAGTTGAGCAGCAGCACTCATGCCGCCCTTGCACCGGCAT 121		
Db	1209	GGTCCCGGGCGGGCGCGGAGCGCGCGCGGTGGGCGAGCGCCCTTGCATAGTTGG 1150		
Qy	122	GCGGAACACCGGGCCAGGATCAGTGTGACGACGCGGATCGGCGCTTCGATCTCGACGGT 181		
Db	1149	CCCGATGACCATGTGGCCCGCGTGTGGCGCGGCGCGCTGGCGCGCAGGGTCTCGCGCGC 1090		
Qy	182	CGCGCGCAGCTCGATCTCGTTCGCCCGCGCTCCACACCGAGGGGAACTGGCTCGGTGCGAG 241		
Db	1089	CGGGGGCGCGCTGTGACCACTCTTTGGGCGACTCGGCTTTGACCCCCCGCGCGCGCGCG 1030		
Qy	242	CGGCGAGCCCCAGCGCGGACAGCTCGTCCAGGCTTGGGCTCACGGAACCTCTCGCGGGTGGG 301		
Db	1029	CGCAGCGCGCGCGCGGGCCCGGGCTGGCGCTGACCCACGCGGGGCCCCCGCGCGCGCG 970		
Qy	302	AGCGGAAACCGGCACGGCTCAGAT- ---CCCTGTGAGTGCATTCGGCTCAGTGCCTGGTC 356		
Db	969	GGCGGGCGCGGGGCGCTGCCCGCGGCTGCGCCCGCTTCTGGGGCTTAATCTCTGATCTTGGC 910		
Qy	357	GTCCCCCTTGGCTGGGAGGATAGCGTTTCACACGAGCGGCACCGCGGGCGGGGGGCG 416		
Db	909	ATTAACCTTTGGCGGGCGGGGCGGCGACGCGGGGCGCGCGGTTCGGGGGTCCGGCGGGGTG 850		
Qy	417	GGGGGGCGGTTTACGCCGATCCGCTCGATGACACAGCGGCTGCGGGGTTCGGGGCGGTTCGCGGA 476		
Db	849	GCGGGGCGCGGGCGCAGGTTGCGGGCGCGGGGGGTGGCGGGGCGGGGCGGGCGCGCGG 790		
Qy	477	GATCCGTACCGCCCGGAGACCGCTTCGGCCAGCGCGCGCGG 516		
Db	789	GGGGCGCGGGCGCGGGGCGCGCGCGGGGGCGCAGGGCGCGGGCG 750		

```
RESULT 44
US-10-114-270-97/c
; Sequence 97, Application US/10114270
; Publication No. US20040030110A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Miller, Charles E.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Spathy, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Liu, Ziaohong
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Li, Li
; APPLICANT: Vernet, Corine
; APPLICANT: Zernusen, Bryan D.
; APPLICANT: Gorman, Linda
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Smithson, Glenna
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Gangolli, Escha A.
; APPLICANT: Taupier Jr., Raymond J.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Ji, Weizhen
; APPLICANT: Anderson, David W.
; APPLICANT: Lieté, Mario W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Stone, David J.
; APPLICANT: MacDougall, John R.
; APPLICANT: Rothenberg, Mark E.
; TITLE OF INVENTION: No. US20040030110A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-322C
; CURRENT APPLICATION NUMBER: US/10/114,270
; CURRENT FILING DATE: 2002-11-27
; PRIOR APPLICATION NUMBER: 60/281,086
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,136
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,863
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/281,906
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/282,020
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/282,930
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/282,934
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/283,512
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/283,710
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/284,234
; PRIOR FILING DATE: 2001-04-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 470
; SEQ ID NO 97
; LENGTH: 5571
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS:
; LOCATION: (67)..(5494)
US-10-114-270-97
Query Match 9.1%; Score 47.2; DB 13; Length 5571;
Best Local Similarity 47.1%; Pred.No. 0.0094;
Matches 178; Conservative 0; Mismatches 198; Indels 2; Gaps 1;
```

```
QY 8 CACGAACGACGGGTGTCGCCCCCGCGCCATCAGAACTGCCACTCCCGCGGGGTGAC 67
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
390 CACGGGGCGCGCGGGTCCAGGGCGCAGGCACCTTCGTCTGTCAGAGTCCGGGATGCCGAC 331
QY 68 CAGGTGCACCAAGGTGGAGTTGAGCAGCAGCTCATCGC--CGCTCGCGCGGCATGCCG 125
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
330 GCGCAGCAGCAGCGCGCTGGCCCGGGGCGCTCATCCCCCGCGGGGCCCGGGCGGG 271
QY 126 AAACACCGGGCCAGGATCAGTGCAGCAGCGCGATGCGCGCTCGATCTCGACGGTCCGC 185
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
270 GCCGCTCGCGCGGGGAAAGAGCGCCCGGGGACTCCGCGCGCGCGCCCGCGCCCC 211
QY 186 GCGAGCTCGATCTCGTCGCGCGCGCTCCACACAGGGGAACTGGCTCGGTGGCAGCGGC 245
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
210 CTTCCCCCGCGGAGCCCGCGCGCTCGCGGGAGAGGGCGGGGGGGGGGGGGGGGG 151
QY 246 AGCCCCAGCGGGACAGCTCGTCCAGGCTGGCGCTCAGCGACCTCTCCGCGGGTCCGGAGCG 305
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
150 GCGCGGGCGCGCGGGGACAGCGGCTCCGGGGGCTCCGAGGGGCGCGCGCGCGGC 91
QY 306 GAAACGGCGCAGGCTCAGATCCCTGTGTCAGTGGCATCGGCTCAGTCCGCTCGTCCCCCTT 365
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
90 GCGCGCGCGCGGCTCAGTGCATCGGCCCGCGCTCAGCTCGCGCGCTGCTGCTTCCCGCGG 31
QY 366 GGCCTGGGAGGATAGCGG 383
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30 GCGCGGGCGCGGGGGG 13
```

```
RESULT 45
US-10-282-122A-30147/c
; Sequence 30147, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
```







; PRIOR APPLICATION NUMBER: US 60/224,729  
; PRIOR FILING DATE: 2000-08-11  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PERL Program  
; SEQ ID NO 38  
; LENGTH: 7789  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No: 7477141CB1  
US-10-333-314-38

Query Match 9.0%; Score 46.8; DB 13; Length 7789;  
Best Local Similarity 50.0%; Pred. No. 0.011;  
Matches 117; Conservative 0; Mismatches 117; Indels 0; Gaps 0;  
Qy 29 CCGGCCGCCCTCAGAACTGCCACTCCGGCGGGGTGACCAAGTCCGACAGGTGGAGTT 89  
Db 4589 CCGTCCGAACAGCGGCGTGTGCGGCCGACGCGCCCGAGTCTCTCGTGTGCCACT 4530  
Qy 89 GAGCAGCCAGCTCATCGCGCTCGCGCGCATGCCGAAACACCGGSCCAGGATCAGTG 148  
Db 4529 GCGTGCATTCGGCTGACAGCCGCTCAGGGTGAAGCTCAGCGCCTGCGCATGCCAG 4470  
Qy 149 CAGCAGCGCGATGCGCGCTCGATCTCAGCGTCCGCGCGAGCTCGATCTCGTCCGCCG 208  
Db 4469 CACCGGGAGCCCGCGCGAGCTCCGCGCTCCGAGCTCTCTCCGTCCCGCGCGGGC 4410  
Qy 209 CTCCACACACAGGGGAACTGGCTCGGTGGCAGCGCGAGCCCGAGCGGGACAG 262  
Db 4409 CTCCAGCGGTGGCGCTCCGCGGAGGGTCCCGCGAGCGCGCTGGGACAG 4356

Search completed: June 27, 2004, 21:03:00  
Job time : 374.196 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 18:07:12 ; Search time 2109.57 Seconds  
(without alignments)  
7360.905 Million cell updates/sec

Title: US-09-758-759-1\_COPY\_109000\_109519  
Perfect score: 520  
Sequence: 1 ggtgcagcagcagcagcgg.....ggccagcgccggcgatcc 520

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_estum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: gb\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_sesl:\*  
29: gb\_ses2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	76	14.6	925	29	CNS0091P
C 2	75	14.4	982	13	BX415111
C 3	68.8	13.2	925	29	CNS0091P
C 4	67.6	13.0	924	13	BX442207

RESULT 1  
CNS0091P/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RFL-98 library from Drosophila melanogaster (fruit fly); genomic survey sequence.  
AL053013  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 925)  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage :  
Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

ALIGNMENTS

5	65	12.5	382	13	BX415111
6	64.2	12.3	932	29	CNS0072Q
7	64	12.3	1288	13	BQ578719
8	63.4	12.2	1153	29	AG065289
9	62.2	12.0	1124	29	AG041123
C 10	61.4	11.8	791	29	CG243505
C 11	61.4	11.8	932	29	CNS0072Q
12	60.2	11.6	844	29	AG083352
C 13	59.6	11.5	1409	12	BG850730
C 14	59.2	11.4	924	13	BX442207
15	58.8	11.3	767	29	AG076679
16	58.8	11.3	835	14	CB904973
17	58.8	11.3	835	14	CF876206
C 18	58.6	11.3	935	29	CNS0065X
C 19	58.6	11.3	1160	13	BUS37937
C 20	58.4	11.2	935	29	CNS0065X
C 21	58	11.2	1015	28	BZ569259
22	58	11.2	1221	29	AG058113
23	57.8	11.1	693	29	AG166257
24	57.8	11.1	878	12	BG785566
C 25	57.8	11.1	1189	29	AG032118
C 26	57.4	11.0	1387	10	AW731151
27	57	11.0	1201	13	BX405071
28	57	11.0	1785	28	CC219595
29	56.8	10.9	1085	29	AG050224
C 30	56.8	10.9	1610	28	BZ569386
C 31	56.6	10.9	483	9	AJ442101
C 32	56.6	10.9	593	29	CG465713
C 33	56.4	10.8	976	29	AG041117
C 34	56.2	10.8	978	29	AG042900
35	56	10.8	1203	29	CNS015Y4
C 36	55.8	10.7	1449	12	BM912953
37	55.8	10.7	851	29	AG171188
38	55.8	10.7	1273	12	BM562099
39	55.6	10.7	1094	14	CK161715
C 40	55.6	10.7	1203	29	CNS015Y4
41	55.6	10.7	1348	29	CG752544
42	55.4	10.7	818	14	CK200413
43	55.4	10.7	1086	29	AG077581
C 44	55.4	10.7	1232	29	AG072425
C 45	55.2	10.6	608	28	BZ639482
C 46	55.2	10.6	633	29	CG623358
C 47	55.2	10.6	912	29	CG328665
C 48	55.2	10.6	1273	12	BM562099
C 49	55	10.6	734	13	BU400981
C 50	55	10.6	1285	29	AG058001



BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Oosawa and  
Aaron Mammoxer in Peter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCL-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source  
1..925  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR19D16"  
/clone\_lib="RPCL-98"  
/note="end : TET3"

## ORIGIN

Query Match 13.2%; Score 68.8; DB 29; Length 925;  
Best Local Similarity 15.1%; Pred. NO. 0.21;  
Matches 54; Conservative 178; Mismatches 124; Indels 1; Gaps 1;  
QY 160 TGGCGCCCTCGATCTCGAGGTGGCGGCGAGCTCGATCTGCGCCCGGCTCCACCA 219  
DB 539 KKGSTTTBGGSTTTTSSGSGYKGGSGSGBSGCCSCSSSSSCCBCCCCSS 598  
QY 220 GGGGAAACTGGCTCGGTGGCGAGCGGCGAGCCCGCGGACAGCTCGTCCAGGCTGGCT 279  
DB 599 YCCSSBSBSKCSSTBSGSCCCSKSVGTCSSSSSCSSSSSTSSSTSKSS 658  
QY 280 CACGACCTTCGCGGTGGGAGCGGAAACGCGACGCTAGATCCCTGTAGTCGCA 339  
DB 659 SGSSSSSSSYTTSKTSASGSGWSAGGSGSTGTSSSSSSSTSTSSSVSGSKS 718  
QY 340 TCGGCTCAGTCCGCGTCCCTTGGCTGGAGGATAGCGGTTCAGACGAGCGCA 399  
DB 719 TBSGSSBSGSSSSSTSSBBSCTSTSSSSSYSSSTCSCTCCCSYSSSTSSSS 778  
QY 400 CACGCGGCGCGGGGCGGGCGGTTCAGCGATCCGCTCGATGACGAGCGGTGCGG 459  
DB 779 TSW-GSTSSSSSVGTSSSSSDSTSTCCSCCYMCTCCSTYBMCVTSTSCGSSSSGK 837  
QY 460 GTCGGGCGGTGCGCGAGATCCGTACCGCCGAGCCGCTCGSCCAGCGCCCGGG 516  
DB 838 GVTKCGCGGSSSTNGMGTSSACSSSSSSSSSSSKSSSSSVSSSSSS 894

RESULT 4  
BX442207  
LOCUS  
DEFINITION  
BX442207 Homo sapiens FETAL BRAIN Homo sapiens cdna clone  
CS0D202YD08 5-PRIME, mRNA sequence.  
ACCESSION  
BX442207  
VERSION  
BX442207.1 GI:30786042  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1. (bases 1 to 924)  
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL  
Contact: Genoscope

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :  
<http://fulllength.invitrogen.com/InvitrogenCorporation1600>  
Faraday Avenue Genoscope sequence ID : CS0D202YD04QF1.

## FEATURES

source  
1..924  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D202YD08"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

## ORIGIN

Query Match 13.0%; Score 67.6; DB 13; Length 924;  
Best Local Similarity 37.5%; Pred. No. 0.33; 247; Indels 4; Gaps 2;  
Matches 192; Conservative 69; Mismatches 247; Indels 4; Gaps 2;  
QY 4 GCACACGACGACGCGGTGGTGGCCCGCGGCCCATCACGAACTGCCATCCGCGGG 63  
DB 408 GGNAGADNAATTTNTTGTGGCGSGGCGCGGGCAGCCGGGCGCGGGCGCGGG 467  
QY 64 TGACAGGTTCAGCAGGTGGAGTTGAGCAGCCAGCTATCGCCGCTCGCGCGGATGC 123  
DB 468 CCGCCCGGCGV 527  
QY 124 CGAAACACCGGCGCAGG---ATCACGTGCAGCAGCGGATCGCGGCTCGATCTCGACGG 180  
DB 528 CNSNNRSCNGSGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 587  
QY 181 TCGCGCGCAGCTCGATCTGTCGCCCGCTCCACACACGAGGGAATGCTCGTGGCA 240  
DB 588 CCG 647  
QY 241 CCGCGCAGCG 300  
DB 648 GCG 706  
QY 301 GAGCGGAAACGCGCAGCGCTCAGATCCCTGTAGTCGATCGGCTCAGTCCGCTCGTCC 360  
DB 707 SGCGCGGCGCMACGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 766  
QY 361 CCCTTGGCTGGGAGGATAGCGGTTCAGCAGCGCGCACACGCGCGCGCGCGCG 420  
DB 767 GCGGGGCG 826  
QY 421 GCGGTTGAGCGATCCGCTCGATGACAGCGGCTGCGGGGTGCGGGCGGTGCGGAGATC 480  
DB 827 GCG 886  
QY 481 CGTACCG 512  
DB 887 CGCCCG 918

RESULT 5  
BX415111  
LOCUS  
DEFINITION  
BX415111 Homo sapiens THYMUS Homo sapiens cdna clone CS0CAF004YI20  
5-PRIME, mRNA sequence.  
ACCESSION  
BX415111  
VERSION  
BX415111.1 GI:30765470  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)

```

ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS        Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
TITLE          Full-length cDNA libraries and normalization
               Unpublished (2001)
COMMENT        Contact: Genoscope
               Genoscope - Centre National de Sequencage
               BP 191 91006 EVRY cedex - France
               Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
               Library was constructed by Life Technologies, a division of
               Invitrogen. Contact : Feng Liang Email : fliang@litech.com URL :
               http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
               Faraday Avenue Genoscope sequence ID : CSOCAP004BE10QPI.
FEATURES       source
               1..982
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="CSOCAP004Y120"
               /tissue_type="THYMUS"
               /clone_lib="Homo sapiens THYMUS"
               /note="Vector; pCMVSPORT 6; 1st strand cDNA was primed
               with a NotI-oligo(dT) primer. Five prime end enriched,
               double-strand cDNA was digested with Not I and cloned into
               the Not I and EcoRV sites of the pCMVSPORT 6 vector.
               Library was not normalized."
ORIGIN
Query Match      12.5%; Score 65; DB 13; Length 982;
Best Local Similarity 29.5%; Pred. No. 0.85;
Matches 129; Conservative 102; Mismatches 206; Indels 0; Gaps 0;
Qy 79 GGTGGAGTTAGCAGCAGCTATCCCGCTCGCGCGGATCGCGAACAACACGGGCCA 138
Db 122 GGGGGGGSSGGVGGVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 181
Qy 139 GATCAGCTGACGACGCGGATGCGGCTCGATCTCGACGCTCGCGCGCAGCTCGATCT 198
Db 182 CCSSCGKCCCGKCCCCCCCCCGGGGGGCCCCSSSSSSSSSSSSSSSSSSSSSSSS 241
Qy 199 GTCGCCCGGCTCCACACAGAGGGAACCTGGCTCGGTGAGCAGCGGACCGCCAGCGGG 258
Db 242 CCCCCCCCCCGCCCCCGGGGCCCCCCCCSSSSSSSSSSSSSSSSSSSSSSSSSSSG 301
Qy 259 ACAGTCGTGACGGTGGCGTCACGGACCTCTCGCGGCTCGGGAGCGGAACGCGCACGG 318
Db 302 GSGGGGGCGSSSCSGSGGGSSSSSSSGSGSCCCCGGGGGSGSGSSSGSCSCSCSG 361
Qy 319 CTCAGATCCCTGTTCAGTCGATCGGCTCAGTCGCGGTCTGTCCTTGGCTTGGAGGAT 378
Db 362 GGGSSSSSCGSSSCSSSGGGSSSSSGGGSSSGGGSSSMSSSSGGCCCCCGCCSCGG 421
Qy 379 AGCGTTTCACACAGCGGACCAACGCGGCGGGGGGGGGGGGGGTTTCAGCCGATCCG 438
Db 422 GSSSGSGSGSGSGSGGGGSSCGCGCGSGGGGGGGGGGGGGGCGGGGGGGGGGGGG 481
Qy 439 TCAGTACACAGCGGCTCGGGGTCGGGCGGTTCGGCGGATTCGTTACCGCCGACCGCC 498
Db 482 GSCGGGGGGGCGSSCGSSGGGGGGGGGGGGSSCGSCCGGGCCCGGGGGGGGGGGGG 541
Qy 499 TCGGCGACGCGCGCGG 515
Db 542 CCCCCCGGCCCCCAG 558

```

```

RESULT 6
CNS0072Q      932 bp DNA linear GSS 03-JUN-1999
LOCUS         Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION    BAC14B09 of RPCI-98 library from Drosophila melanogaster (fruit
               fly), genomic survey sequence.

```

```

AL066742      GI:4945205
AL066742.1    GSS.
SOURCE        Drosophila melanogaster (fruit fly)
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 932)
AUTHORS       Direct Submission
TITLE         Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL       BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT       - Web : www.genoscope.cns.fr)
               Determination of this BAC-end sequence was carried out as part of a
               collaboration with the Berkeley Drosophila Genome Project (BDGP).
               The BDGP is constructing a physical map of the Drosophila
               melanogaster genome using these BACs. For further information
               please see http://www.fruitfly.org The BDGP Drosophila
               melanogaster BAC library was prepared by Kazutoyo Oseegawa and
               Aaron Mammoss in Pieter de Jong's laboratory in the Department of
               Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
               NY. The library is named RPCI-98 and was constructed by partial,
               EcoRI digestion of Drosophila DNA provided by the BDGP from the
               isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
               P1 and EST libraries. A more detailed description of the library
               and how to order individual BAC clones, the entire library, or
               filters for hybridization from the BACPAC Resource Center can be
               found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
               Location/Qualifiers
               1..932
               /organism="Drosophila melanogaster"
               /mol_type="genomic DNA"
               /db_xref="taxon:7227"
               /clone="BAC14B09"
               /clone_lib="RPCI-98"
               /note="end : T7"
ORIGIN
Query Match      12.3%; Score 64.2; DB 29; Length 932;
Best Local Similarity 30.9%; Pred. No. 1.1;
Matches 109; Conservative 86; Mismatches 158; Indels 0; Gaps 0;
Qy 161 GCGCGCTTCGATCTCGACGCTCGCGCGGAGTCTGATCTCGCGCGGCTCCACACGAG 220
Db 555 GSKSGSGSGSGSBGCGGGGTGSGGGGGGGBTKYTBKGGGGGGSGSGSGSGSGSG 614
Qy 221 GGGAACTGGCTCGGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTC 280
Db 615 GGGGCGSGSGSGCGCGGCGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 674
Qy 281 ACGGACCTTTCGCGGCTCGGGAGCGGAAACGCGACGCTCAGATCCCTGTTCAGTCGAT 340
Db 675 GCGGCGSGSGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 734
Qy 341 CGGCTCAGTTCGCTCGTCCCTTGGCTTGGAGTAGCGGTTTCAGACGAGCGGCGAC 400
Db 735 CCCCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 794
Qy 401 CACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 460
Db 795 CGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 854
Qy 461 TCGGGGCGGCTCGCGAGATCCGTACGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 513
Db 855 SSGCGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 907

```

```

RESULT 7
CNS0072Q      1288 bp mRNA linear EST 15-JUL-2002
LOCUS         BO678719
DEFINITION    AGENCOURT_8199257 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6261563
               5', mRNA sequence.

```

```
ACCESSION BQ678719
VERSION BQ678719.1 GI:21791398
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1288)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/BTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2422 row: h column: 12
High quality sequence stop: 171.
FEATURES
    source
    1..1288
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:6261563"
        /tissue_type="melanotic melanoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /clone_lib="NIH_MGC_112"
        /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
        EcoRI; cDNA made by oligo-dT priming. Directionally cloned
        into EcoRI/XhoI sites using the following 5' adaptor:
        GGCACGAG(G). Library constructed by Ling Hong in the
        laboratory of Gerald M. Rubin (University of California,
        Berkeley) using Zap-cDNA synthesis kit (Stratagene) and
        Superscript II RT (Life Technologies). Note: this is a
        NIH_MGC Library."
ORIGIN
    Query Match 12.3%; Score 64; DB 13; Length 1288;
    Best Local Similarity 42.8%; Pred. No. 1.3; Mismatches 0; Indels 0; Gaps 0;
    Matches 205; Conservative 0; Mismatches 274; Indels 0; Gaps 0;
QY 22 GGTCCGCCCGCGCGCCATCGAAGTCCACTCGCGGGGTGACCAAGTTCGACCAAGT 81
DB 498 GGGCGCGCGCGCGCCCGCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547
QY 82 GGGAGTTGAGCAGCAGTCTATCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 141
DB 548 GCGGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 607
QY 142 TCACGTGACGACGCGCGATGCGCGCGCTCGATCTCGACGTCGCGCGCGCGCGCG 201
DB 608 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 667
QY 202 CGCCCGGCTCCACACAGGGGAACTGGCTGGTGGCAGCGGACGCCAGCGCGGACA 261
DB 668 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 727
QY 262 GCTCGTCAGAGTGGCGTCAAGCACTCTCGCGGTTCGGAGCGGAAACGCGCAGG 321
DB 728 NCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787
QY 322 AGATCCCTGTGAGTCAGTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 381
DB 788 CNGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 847
QY 382 GTTTCACGACGAGCGGACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 441
DB 848 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 907

QY 442 ATGACACGCGCGTCCGGGGTCCGGGGTCCGGGGTCCGGGGTCCGGGGTCCGGGGT 500
DB 908 CMC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966

RESULT 8
AG065289 1153 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-054K01.F, genomic survey sequence.
AG065289
ACCESSION AG065289
VERSION AG065289.1 GI:16617091
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1153)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
[E-mail: chimpanse@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170]
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
    source
    1..1153
        /organism="Pan troglodytes"
        /mol_type="genomic DNA"
        /db_xref="taxon:9598"
        /clone="PTB-054K01.F"
        /sex="male"
        /cell_type="lymphoblast"
        /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
    Query Match 12.2%; Score 63.4; DB 29; Length 1153;
    Best Local Similarity 46.3%; Pred. No. 1.6; Mismatches 277; Indels 1; Gaps 1;
    Matches 240; Conservative 0; Mismatches 277; Indels 1; Gaps 1;
QY 4 GCAGCAGCAACGACCGGTGTGTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 63
DB 211 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 270
QY 64 TGACACAGGTTCGACGAGTGGAGTTGAGCAGCGAGTCTATCGCGCGCTCGCGCGGATGC 123
DB 271 CGGCGACGACGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 330
QY 124 CGAAACACCGCGCGAGGATCACTGTCAGCACGCGCGCGCGCGCGCGCGCGCG 183
DB 331 CGGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
QY 184 GCGCGAGTTCGATCTCGTTCGCCCGGTCTCCACACAGGGGAACTGGCTCGGTGGCAGCG 243
DB 391 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 450
QY 244 GCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 303
DB 451 CGGCGAGCGCGCGAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 510
```



Matches 232; Conservative 0; Mismatches 266; Indels 1; Gaps 1;

QY 22 GGTGCGCCCGCGCCATCAGAACTGCACTCCGCGGGGTGACACAGGTGCGACAGGT 81

Db 503 GTCTGCGCCCTGTGCTTGTGACACGAGGCTCTGTCAGGCTCTTGGACCGCGCGC 444

QY 82 GGGAGTTGAGCAGCAGCTATCGCGCGCTGCGCGGATCGCGAAACACCGGCGCAGGA 141

Db 443 CCGCCCGCGCGCGCGCGCGAGCGCGCGCGCGCGCGATCGGAACTCGCGGCTGCGGA 384

QY 142 TCACGTGACGACGCGCGATGCGCGCTCGATCTCAGCGTGGCGCGAGCTCGATCTGCT 201

Db 383 CCAGAGCAGAGTTGTCGCTGAACCGAGTGCCTCCAGCGCGCGCGCTCTCTCAGGG 324

QY 202 CCGCCGCGTCCACACGAGGGGAAATGCTCGTGGCAGCGGACAGCCCGCGCGACA 261

Db 323 TGAGACCGACGCTCGCGGTCAGGAGCGCGCGCTCGCTCGCGCTCTCTCAGAGCTCC 264

QY 262 GCTGCTCAGGCTGCGCTCAGGACCTCTCGCGGTCGGGAGCGGAGAAACCGGACGGTCT 321

Db 263 TCTCCACTGCTCGCGGCGCCAGTCTCTCGCGCCCTCGCGCGGAGTGGCGCCCAACACG 204

QY 322 AGATCCCTGTAGTCCGATCGCTCAGTCCGCTGCTGCTCCCTTGGCTGGGAGATAGC 381

Db 203 GGACAGCTCCACG 144

QY 382 GTTTCAGCAGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 441

Db 143 CTTCCACGAGGACCACTCCAGCGCGCTCCCGCGCTCGCGCTCGATCTT-GCTGCGCGTG 85

QY 442 ATGACAGCGCTGCGGGGTGCGGGGTGCGCGGAGTCCGTAACCGCGCGCGCGCGCG 501

Db 84 AAGATCGGAGCTGCGGGCGCTCGGAAACGACGCTTCCAGCGCGCGCGCTTCCGCGCC 25

QY 502 GCAGGCGCGCGGATCC 520

Db 24 GCGCTCTCGCGCGCGCC 6

RESULT 11

CNS00720/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Drosophila melanogaster genome survey sequence T7 end of BAC # BAC14809 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL066742

AL066742.1 GI:4945205

GSS

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 932)

Genoscope

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

Source

Location/Qualifiers

1..932

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BAC14809"

/clone\_lib="RPCI-98"

/notes="end : T7"

## ORIGIN

Query Match 11.8%; Score 61.4; DB 29; Length 932;

Best Local Similarity 32.0%; Pred. No. 3.1;

Matches 108; Conservative 76; Mismatches 153; Indels 0; Gaps 0;

QY 169 CGATCTCAGCTCGCGCGCGCTCGATCTCGTCGCGCGCTCCACACAGCGGAACT 228

Db 929 CG 870

QY 229 GGTCTGCTGCG 288

Db 869 GCG 810

QY 289 CTGCGCGCTGCGGAGCGGAAACGCGCGCTCAGATCCCTGTCAGTTCGATCGGCTCAG 348

Db 809 SSSCG 750

QY 349 TGGCGCTGCTGCG 408

Db 749 CGSGCG 690

QY 409 GCG 468

Db 689 GCG 630

QY 469 GTGCGCGAGATCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 505

Db 629 CG 593

## RESULT 12

AG083352

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Pan troglodytes DNA, clone: PTB-080M08.R, genomic survey sequence.

AG083352

AG083352.1 GI:16635154

GSS

Pan troglodytes (chimpanzee)

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

1

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 844)

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chmpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : pKS145

R.Site 1 : SacI







```

QY 496 GCTTCGGCCACGCGCGCGG 516
Db 567 CGGCGCGCGCGCGCGG 587

RESULT 16
CB904973 835 bp mRNA linear EST 02-JUL-2003
LOCUS trico40xn11 T.reesei mycelial culture, Version 3 april Hypocrea
DEFINITION jecorina cDNA clone trico40xn11, mRNA sequence.
ACCESSION CB904973
VERSION CB904973.1 GI:301119631
SOURCE EST.
ORGANISM Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 835)
Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
Dunn-Coleman,N.S., Godegebuur,F., Houfek,T.D., England,G.J.,
Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)
JOURNAL 22803114
MEDLINE 12789320
PUBMED
COMMENT Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.
Location/Qualifiers
1. .835
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trico40xn11"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/note="Vector: pREP3y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Query Match 11.3%; Score 58.8; DB 14; Length 835;
Best Local Similarity 46.7%; Pred. No. 7.9;
Matches 243; Conservative 0; Mismatches 273; Indels 4; Gaps 2;

QY 2 GTGCAGCAGCAACGCGGTGGTCGCGCGCGCCCATCACGAATGCCACTCCGCGG 61
Db 299 GGGCGCGGACACGAGACCCCAACACGCCCCCGGAGAGAGCGCGNCGACAAG 358
QY 62 GGTGACACAGTCGACCGAGTGGAGTTGAGCAGCGCCTCATCGCCGCTGCGCGGCAT 121
Db 359 CCGCGCGCGACGACGCGCGCGCGCGCACGCCANCGACNCGCGCGCGCGCGAC 418
QY 122 GCCGAACACCGCGCCAGGATCAGTCGACGACGCGCATGCGGCTTCGACGGT 181
Db 419 GGCACGAGCGCGCGCCAGCAGACGACGACGACGACGACGACGACGACGAC 478
QY 182 CGGCGCGAGTCGATCTCGTCCCGCGCTCCACACAGGGGAACTGGTCGTCGCGAG 241
Db 479 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 538
QY 242 CGGCGAGCGCGCGGACAGCTGTCAGGCTGCGCTCAGGACCTCTCGCGGTCGG 301
Db 539 AGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 595

```

```

QY 302 AGCGGAAACGGCGACGCGCTCAGATCCCTGTGTCGATCGCGCTCAGTCCGCTGTCGCC 361
Db 596 CGCGAGGACCGCGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 655
QY 362 CTTGCGCTGGAGATAGCGTTTACGACGAGGCGGACACGCGGCGCGCGCGCGCGG- 420
Db 656 ACGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 715
QY 421 GGGGTTCAGCGATCCGCTCGATGACACGCGGCTCGCGGCGTCCGGCGCGTTCGGCGGATC 480
Db 716 GGGCGCGAGCGGGGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 775
QY 481 CGTACCGCGCGGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATCC 520
Db 776 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 815

RESULT 17
CF876206 835 bp mRNA linear EST 31-OCT-2003
LOCUS trico40xn11.b1 T.reesei mycelial culture, Version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone trico40xn11, mRNA sequence.
ACCESSION CF876206
VERSION CF876206.1 GI:38130888
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 835)
Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
Dean,R.A.
Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
Unpublished (2003)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: LT-F1 primer.
Location/Qualifiers
1. .835
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trico40xn11"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/note="Vector: pREP3y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Query Match 11.3%; Score 58.8; DB 14; Length 835;
Best Local Similarity 46.7%; Pred. No. 7.9;
Matches 243; Conservative 0; Mismatches 273; Indels 4; Gaps 2;

QY 2 GTGCAGCAGCAACGCGGTGGTCGCGCGCGCCCATCACGAATGCCACTCCGCGG 61
Db 299 GGGCGCGGACACGAGACCCCAACACGCCCCCGGAGAGAGCGCGNCGACAAG 358
QY 62 GGTGACACAGTCGACCGAGTGGAGTTGAGCAGCGCCTCATCGCCGCTGCGCGGCAT 121
Db 359 CCGCGCGCGACGACGCGCGCGCGCGCACGCCANCGACNCGCGCGCGCGCGAC 418
QY 122 GCCGAACACCGCGCCAGGATCAGTCGACGACGCGCATGCGGCTTCGACGGT 181
Db 419 GGCACGAGCGCGCGCCAGCAGACGACGACGACGACGACGACGACGACGAC 478
QY 182 CGGCGCGAGTCGATCTCGTCCCGCGCTCCACACAGGGGAACTGGTCGTCGCGAG 241
Db 479 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 538
QY 242 CGGCGAGCGCGCGGACAGCTGTCAGGCTGCGCTCAGGACCTCTCGCGGTCGG 301
Db 539 AGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 595

```



of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

```

ORIGIN
Query Match      11.3%; Score 58.6; DB 13; Length 1160;
Best Local Similarity 45.9%; Pred. No. 9;
Matches 162; Conservative 0; Mismatches 187; Indels 4; Gaps 1;

QY 129 CACGGGCCAGGATCAGTGCAGACGCGGATCGGCGCTTCGATCTCGAGCGTTCGCGCCG 188
Db 1116 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1057
QY 189 AGTCTGATCTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 248
Db 1056 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 997
QY 249 CCGAGCGGACAGCTCTGTCAGGCTGCTCAGGACCTCTCTCGCGGTCGCGAGCGGAA 308
Db 996 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 937
QY 309 ACGCG----CACGGCTCAGATCCCTGTCACTCGGCTCAGTCTCGGTCGTCAGCGCAGC 364
Db 936 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877
QY 365 TGGCTCTGAGGATAGCGGTTTACGACGAGCGGACACAGCGCGCGCGCGCGCGCGCGCG 424
Db 876 GCGCGCGGNNNGGGGGGGGNNNGGGGGGNNNGGGGGGNNNGGGGGGNNNGGGGGGGG 817
QY 425 TTACAGCGCATCCGCTCAGTCAACAGCGCGCTCGCGGGTCGCGGGTCGCGCGAG 477
Db 816 GCGGGGGGNNNGGGGGGNNNGGGGGGNNNGGGGGGNNNGGGGGGNNNGGGGGG 764

```

```

RESULT 20
CNS006XXK
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence, 935 bp DNA linear GSS 03-JUN-1999
  BAC14N09 of RPII-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION
  AL066051.1 GI:4945019
VERSION
  GSS.
KEYWORDS
  Drosophila melanogaster (fruit fly)
SOURCE
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 935)
ORGANISM
  Genoscope.
  Direct Submission
  Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
COMMENT
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazutoyo Osoegawa and
  Aaron Mamooser in Pieter de Jong's laboratory in the Department of
  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
  NY. The library is named RPII-98 and was constructed by partial
  EcoRI digestion of Drosophila DNA provided by the BDGP from the
  isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
  P1 and EST libraries. A more detailed description of the library
  and how to order individual BAC clones, the entire library, or
  filters for hybridization from the BACPAC Resource Center can be
  found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
  Location/Qualifiers
    1..935
    /organism="Drosophila melanogaster"
    /mol_type="genomic DNA"

```

/db\_xref="taxon:7227"  
/clone="BAC14N09"  
/clone\_lib="RPII-98"  
/note="end : 17"

```

ORIGIN
Query Match      11.2%; Score 58.4; DB 29; Length 935;
Best Local Similarity 28.8%; Pred. No. 9.3;
Matches 121; Conservative 107; Mismatches 191; Indels 1; Gaps 1;

QY 94 GCAGCTCATCGCCCTCGGCGCGGATCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 213
Db 508 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 567
QY 154 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 213
Db 568 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 627
QY 214 ACACGAGGGGAACTGCGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 273
Db 628 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 687
QY 274 TGGCGTCAAGCACTCTCTCGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 333
Db 688 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 746
QY 334 CTCGATCGGTCATGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCG 393
Db 747 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 806
QY 394 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 453
Db 807 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 866
QY 454 TCGGCGGTCGCGGCGGTCGCGGCGGTCGCGGCGGTCGCGGCGGTCGCGGCGGTCGCGGCG 513
Db 867 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 926

RESULT 21
BZ569259/c
LOCUS
DEFINITION
  BZ569259
  clone pac2-164_8319, genomic survey sequence.
ACCESSION
  BZ569259.1 GI:27203589
VERSION
  GSS.
KEYWORDS
  Pseudomonas aeruginosa
  Pseudomonas aeruginosa
  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
  Pseudomonadaceae; Pseudomonas.
REFERENCE
  1 (bases 1 to 1015)
  Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
  Burns, J.L., Kaul, R. and Olsen, M.V.
  Whole-Genome-Sequence variation among multiple isolates of
  Pseudomonas aeruginosa library
  J. Bacteriol. (2002) In press
  Contact: Chris K. Raymond
  Genome Center
  University of Washington
  Box 352145, Seattle, WA 98105-2145, USA
  Tel: 2062216954
  Fax: 2066857244
  Email: craymond@u.washington.edu
  Class: shotgun.
  Location/Qualifiers
    1..1015
    /organism="Pseudomonas aeruginosa"
    /mol_type="genomic DNA"
    /strain="2-164"
    /db_xref="taxon:287"
    /clone="pac2-164_8319"
    /clone_lib="pac2-164"

```





[illegible]







Class: shotgun.  
Location/Qualifiers  
1. 1610

## FEATURES

source  
1. 1610  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pacs2-164.8365"  
/clone\_lib="pacs2-164"  
/note="clinical isolate 2-164 Whole genomic shotgun library."

## ORIGIN

Query Match 10.9%; Score 56.8; DB 28; Length 1610;  
Best Local Similarity 46.1%; Pred. No. 18;  
Matches 224; Conservative 0; Mismatches 254; Indels 8; Gaps 2;  
QY 27 CCCCGCGCGCCATCAGCACTGCACCTCCGCGGGGTGACACAGTGCACAGTGGAG 86  
DB 1235 CG 1176  
QY 87 TTGAGCAGCCAGCTCATCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 146  
DB 1175 CCTTNGCGGCG 1116  
QY 147 TGCAGCAGCGCGGATGCGCGCGCTCGATCTCGACGCTGCGCGCGCGCGCGCG 206  
DB 1115 CG 1056  
QY 207 GGCTCCACACACAGGCGAAATGCTCTGCTGCGCGCGCGCGCGCGCGCGCGCG 266  
DB 1055 GCG 996  
QY 267 TCCAGGTGGCGTCAACGACCTCTCGCGGTCGCGGAGCGGAAACGCGCAGGTCAGATC 326  
DB 995 GCG 937  
QY 327 CTTCTCAGTCAGTCAGTCTCAGTCTCAGTCTCAGTCTCAGTCTCAGTCTCAGTCTC 386  
DB 936 GCGC-----CGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 884  
QY 387 ACAGCAGCGCGCACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 446  
DB 883 NGCG 824  
QY 447 CAGCGGCTGCGGGGTGCGGGCGGTGCGGCGAGATCCGTACCGCGCGCGCGCGCGCG 506  
DB 823 CCG 764  
QY 507 GCGCGC 512  
DB 763 GCGCGC 758

RESULT 31  
LOCUS AJ442101/C  
DEFINITION AJ442101 dkf426 Gallus gallus cDNA clone 19m6r1, mRNA sequence.  
ACCESSION AJ442101  
VERSION AJ442101.1 GI:20209322  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 483)  
Buerstedde, J.M.  
Gallus gallus bursal lymphocyte EST  
Unpublished (2002)  
Contact: Buerstedde JM  
Cellular Immunology  
Heinrich-Pette-Institute

Martinistr. 52, 20251 Hamburg, Germany  
Email: URL: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.

## FEATURES

source  
1. 483  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="CB"  
/db\_xref="taxon:9031"  
/clone="19m6r1"  
/tissue\_type="Bursa of Fabricius"  
/cell\_type="Bursal lymphocyte"  
/dev\_stage="2-3 weeks old"  
/clone\_lib="dkfz426"

## ORIGIN

Query Match 10.9%; Score 56.6; DB 9; Length 483;  
Best Local Similarity 49.8%; Pred. No. 16;  
Matches 143; Conservative 0; Mismatches 144; Indels 0; Gaps 0;  
QY 32 GCGCGCCCATCAGCACTGCCACTCGCGGGGTGACACAGTGCACAGTGGAGTTGAG 91  
DB 389 GCGGAGCAGCG 330  
QY 92 CAGCCAGCTCATCTGCGCGCGCTGCGCGCGCATGCCGAAACACCGCGCGCGCGCG 151  
DB 329 CAGAGCAGCGCGCGGTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 270  
QY 152 CAGCGCGATGCGCGCGCTCGATCTCGACGCTGCGCGCGCGCGCGCGCGCGCG 211  
DB 269 CAGCG 210  
QY 212 CCACACACAGGCGAACTGGCTCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 271  
DB 209 GCG 150  
QY 272 GCTGCGGTGACGAGACTCTCGCGGGTTCGCGAGCGGAAACGCGCGCGCG 318  
DB 149 GCGCTCGCGCGCGCTCTCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103

## RESULT 32

## LOCUS

## CG465713

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

CG465713 593 bp DNA linear GSS 24-SEP-2003  
KRIBB\_2D\_078T7 pBS-NN Library Homo sapiens genomic clone  
KRIBB\_2D\_078 5, genomic survey sequence.  
CG465713  
CG465713.1 GI:35189148  
GSS.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Lee, K.T., Kim, J.H., Hahn, Y.S., Yang, J.O., Chu, M.Y., Kim, H.C.,  
Oh, K.J., Kim, S.S., Yoo, H.S. and Kim, Y.S.  
Rapid isolation of NotI-linked CpG island sequences from RLOS gel  
Unpublished (2003)  
Contact: KIM YS  
Human Genome Research Laboratory  
Korea Research Institute of Bioscience and Biotechnology  
52 Boeun-dong, Yuseong-gu, Daejeon 305-333, Republic of Korea  
Tel: 82-42-879-8110  
Fax: 82-42-879-8119  
Email: yongsung@kribb.re.kr  
This result was produced by the program for Functional Analysis of  
Human Genome funded from Ministry of Science and Technology,  
Republic of Korea. Clones are available from the Center for  
Functional Analysis of Human Genome. See URL:  
<http://21cgenome.kribb.re.kr/> for details or contact:  
yongsung@kribb.re.kr.  
Seq primer: T7 primer  
Class: NotI site  
High quality sequence stop: 593.

```
FEATURES
  source      Location/Qualifiers
1..593
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /clone="KR18B_2D_378"
  /sex="male"
  /tissue type="blood"
  /dev stage="adult"
  /clone lib="pBS-NN Library"
  /note="Vector: pBluescriptKS(+); Site 1: NotI; Site 2:
  NotI; Genomic DNA was digested with NotI and the resulting
  fragments were ligated into ZAPII/NotI vector DNA. After
  the digestion with EcoRV, the mixture was run on 0.8% LMP
  agarose gel, eluted a gel portion over 22 kb and purified
  with beta-agarase I. The purified DNA was digested with
  NotI, run on 1% LMP agarose gel, and eluted with Gel
  Extraction kit the DNA fragments ranging 0.7 to 4 kb,
  which cover on first-dimensional gel comprising the
  central portion of the standard RGS profile. The DNA
  ends. The ligated solution was transformed into DH5a cell
  using electroporation machine."

ORIGIN
  Query Match      10.9%; Score 56.6; DB 29; Length 593;
  Best Local Similarity 48.6%; Pred. No. 16;
  Matches 245; Conservative 0; Mismatches 250; Indels 9; Gaps 3;

QY 3  TGCAGCAGACGACCGGTGTCGCCCGCGGCCCATCAGAACTGCCATCCGCGGG 62
Db 48  TTTCTACCGGGGGGAGCGGAGCCCTCGCGCTCGGCTGCTGAGTGGGTAGC 107
QY 63  GTGACCAAGTGCACAGGTGGAGTTGAGCAGCAGGTATCGCGCTCGCGCGGATG 122
Db 108  GGCAGCGGCCCGCGCGCGCGGTGGGGCAGCAGCGCGCGGAGCCTCGAG 167
QY 123  CCGAAACACCGCGCCAGGATCAGTGCAGCAGCGGATGCGCGCTCGATCTCGACGTC 182
Db 168  TCCAGCAGCTGGGGGNCACAGCGCGCGCTCGCGTGCCTCGCGCGCGC-----C 222
QY 183  GCGCGACGTCGATCTCGTCGCCGCTCCACACAGCGGAAACTGCTCGGTGGCAGC 242
Db 223  TGCCGCTGCCGTGCCGCTGCCAGCCCCCGCGCGCGCAGCCCGCGCAGACGAGCAGC 282
QY 243  GGCACCCCGCGCGGACAGTCTGTCAGGCTGCGTCAAGGCTCAGCGACCTC--TCGCGGGTCGG 300
Db 283  CGGAGCCCGGACCGCGCGGAGGCTATCTGTGTCAAGAGTTCTGTGCCGCGCGCT 342
QY 301  GAGCGAAACCGCAGCGCTCAGATCCCTGTGTCAGTCGATCGGCTCAGTGCCTGTC 360
Db 343  GCGCGGCGCTCCGCGAGGACGAGTTCCACATCAGTGTCTATCAGGTCTGTCGCGCGGTC 402
QY 361  CCCTTGGCTTGGAGGATAGCGGTTTCAGCAGCGGACCAACCGCGCGGGCGGGCGGG 420
Db 403  TGGGGCGCGGGGACAGT--CAGGTGACCGGGCGGGCGGGCGGGGTCGGGGCAGG 460
QY 421  GCGGTTACGCGATCCGCTCGATACACAGCGGCTGCGGGGTGCGGGGTGCGGAGATC 480
Db 461  GCGCGTGGCGGGGGGTGCGGGGCGCGGGGAGCGGGGCGGGGCGGGGCGGGGCGG 520
QY 481  CGTACCGCGCGACCGCTCGGCC 504
Db 521  CGGGACCGCAGCGCGGCAAGCC 544

RESULT 33
AG041117/c
LOCUS      AG041117
DEFINITION Pan troglodytes DNA, clone: PTB-018117.F, genomic survey sequence.
ACCESSION AG041117
VERSION    AG041117.1 GI:16569842
KEYWORDS   GSS.

SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes
MAMMALIA; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE   1
AUTHORS     Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE       BAC end sequences of Library PTB
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 976)
AUTHORS     Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE       Direct Submission
JOURNAL     Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail: chimpes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/;
            Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
PRIMERS
  Sequencing: -21M13
  LIBRARY
  Vector      : pKS145
  R.Site 1    : SacI
  R.Site 2    : SacI
  Location/Qualifiers
1..976
  /organism="Pan troglodytes"
  /mol_type="genomic DNA"
  /db_xref="taxon:9598"
  /clone="PTB-018117.F"
  /sex="male"
  /cell_type="lymphoblast"
  /clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
  Query Match      10.8%; Score 56.4; DB 29; Length 976;
  Best Local Similarity 45.3%; Pred. No. 19;
  Matches 224; Conservative 0; Mismatches 269; Indels 1; Gaps 1;

QY 17  CCGTGGTTCGCCCGCGCGCCCATCAGCAATGCGCATCTCCGCGGGGTGACAGGTGAC 76
Db 844  CCGCGNNGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 785
QY 77  CAGGTGGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCG 136
Db 784  GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 725
QY 137  -CAGGATCAGTGCAGCAGCGGATCGCGCGCTCGATCTCGACGGTTCGCGCGCGCGCA 195
Db 724  GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 665
QY 196  TCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 255
Db 664  CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 605
QY 256  GGGACAGCTCTGTCAGGCTGGCGTCAAGACCTCTCGCGGGTTCGGGAGCGGAAACGCGCA 315
Db 604  GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 545
QY 316  CCGCTCAGATCCCTGTGCTCAGTCCATCGGCTAGTCCGCGGTCTGCTCCCTTGGGCTG 375
Db 544  CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 485
QY 376  GATAGCGGTTCAGCAGCAGCGCGCACCGCGCGGGGGGGCGGGCGGGCGGTTCAGCGCGATC 435
Db 484  CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 425
QY 436  CGTCTGATGACAGCGGCTGCGGGGTTCGGGGTTCGGGCGAGATCCGTACCGCGCGGACC 495
Db 424  GGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 365
```

Qy	496	GCCTCGGCCACGGC	509	
Db	364	CGCCCGCGCGCGC	351	
RESULT 34				
AG042900				
LOCUS				
DEFINITION Pan troglodytes DNA, clone: PTB-021A20.R, genomic survey sequence.				
ACCESSION AG042900				
VERSION AG042900.1 GI:16571625				
KEYWORDS GSS.				
SOURCE				
ORGANISM				
Pan troglodytes (chimpanzee)				
Pan troglodytes				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.				
1				
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,				
Totoki,Y., Watanabe,H. and Sakaki,Y.				
BAC end sequences of library PTB				
Unpublished				
2 (bases 1 to 978)				
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,				
Totoki,Y., Watanabe,H. and Sakaki,Y.				
Direct Submission				
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical				
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);				
1-7-22 Suehiro-chou,Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan				
(E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/;				
Tel:81-45-503-9111, Fax:81-45-503-9170)				
Clones are derived from the chimpanzee BAC library PTB This BAC end				
was generated during the R&D process and may have higher chance of				
clone tracking errors.				
PRIMERS				
Sequencing: M13Rev				
LIBRARY				
Vector : pXS145				
R.Site 1 : SacI				
R.Site 2 : SacI.				
Location/Qualifiers				
1. 978				
/organism="Pan troglodytes"				
/mol_type="genomic DNA"				
/db_xref="taxon:9598"				
/clone="PTB-021A20.R"				
/sex="male"				
/cell_type="lymphoblast"				
/clone_lib="PTB Chimpanzee Male BAC Library"				
ORIGIN				
Query Match				
Best Local Similarity				
Matches 169; Conservative 0; Mismatches 193; Indels 0; Gaps 0;				
155 GCGCATCGCGCCCTCGATCTCGACGGCTCGGCGCAGCTCGATCTCGTCGCCCGGCTCCCA 214				
436 GGGCGCGCGCCCGCGCCCGCGCGCGGGGCGGGGCGCGCGCGCGCGCGCGCGCGCG 495				
Qy				
215 CACGAGGGGAACCTGGCTCGGTGCGACGGCAGGCCACCGCGGACAGCTCGTCAGGCT 274				
Db				
496 GCGCGCGCGGGGGCG 555				
Qy				
275 GCGCTCAGGACCTCTCGCGGGTCGGGAGCGGAAACGCGACGGCTCAGATCCCTGTGAG 334				
Db				
556 CGGGCGCGGGGCG 615				
Qy				
335 TCGCATGGGCTCAGTGGCGGTCGTCCTTCGGCTCGGAGATAGCGGTTCAGACGAG 394				
Db				
616 GCGCGCGGGGCG 675				
Qy				
395 CGGCACCAACGCGCGCGCGCGCGCGGGGGTTCACCGCATCCGCTCGATGACACGCGCT 454				



/sex="male"  
/cell\_type="lymphocytes"  
/clone\_lib="RPGI-43 Chimpanzee Male BAC Library"

ORIGIN

Query Match 10.7%; Score 55.8; DB 29; Length 851;  
Best Local Similarity 43.3%; Pred. No. 24; Mismatches 273; Indels 1; Gaps 1;  
Matches 209; Conservative 0;

QY 14 CGACGGTGGTGGCCCGCCGATGCGGCTCGATCTCGATCTCGATCTCGGCGCGGAGCTC 193  
DB 318 CGGGGCGCCNNCCCNCCGCGGNNCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 377  
QY 74 GACACAGTGGAGTGGAGCAGCAGTCTATCCCGGCTCGCGCGGCGGCGGCGGCGGCGGCGG 133  
DB 378 CGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 437  
QY 134 GGCCAGGATCAGTGCAGCAGCGGATGCGGCTCGATCTCGATCTCGATCTCGGCGCGGAGCTC 193  
DB 438 GCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 497  
QY 194 GATCTCGTCCCGGCTCCACACAGGGGAACTGGCTCGTGGTGGAGCGGCGGCGGCGGCGGCGG 253  
DB 498 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 556  
QY 254 CCGGGACAGTCTCCAGGCTGGCGTACGCGACTCTCGCGGCTCGGAGCGGAAACGCG 313  
DB 557 CCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 616  
QY 314 CACGGCTCAGATCCCTGTCTAGTCCGATCGGCTCAGTCCCGGCTCGTCCCGCTTTGGCTGG 373  
DB 617 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 676  
QY 374 AGATAGCGGTTACGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 433  
DB 677 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 736  
QY 434 TCCTCTCGATGACGAGCGGCTCGGCGGCTCGGCGGCTCGGCGGCTCGGCGGCTCGGCGGCT 493  
DB 737 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 796  
QY 494 CCG 496  
DB 797 CCG 799

RESULT 38  
BM562099  
LOCUS  
DEFINITION BM562099 1273 bp mRNA linear EST 20-FEB-2002  
AGENCOURT\_6597561 NIH\_MGC\_41 Homo sapiens cDNA clone IMAGE:5480748  
5', mRNA sequence.

ACCESSION BM562099  
VERSION BM562099.1 GI:18807897  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1273)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Straubeberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: DCD/DTF

CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov

Plate: LNCM2004 row: b column: 13  
High quality sequence stop: 206.

FEATURES  
source

Location/Qualifiers  
1..1273  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5480748"  
/tissue\_type="amelanotic melanoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_41"  
/notes="Organ: Skin; Vector: pOTB7; Site: 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGACAGAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

ORIGIN

Query Match 10.7%; Score 55.8; DB 12; Length 1273;  
Best Local Similarity 47.5%; Pred. No. 25;  
Matches 214; Conservative 0; Mismatches 232; Indels 5; Gaps 2;

QY 8 CACGAAAGACGGTGGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 67  
DB 594 CG 653  
QY 68 CAGGTGCACAGGTGGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGCGGCGGCGGCGGCG 127  
DB 654 CCGGCG 713  
QY 128 ACACCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 185  
DB 714 GCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 773  
QY 186 CGCAGCTCGATCTCGTCCCGCGGCTCCACACAGCGGCGGCGGCGGCGGCGGCGGCGG 245  
DB 774 CG 833  
QY 246 AGCCCCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 305  
DB 834 CG 890  
QY 306 GAAACGCGCGCGGCTCAGATCCCTGTGATCGATCGGCTCAGTCCGCTCGGCTCGGCTT 365  
DB 891 GCG 950  
QY 366 GCGCTGGGAGGATAGCGGTTTACGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 425  
DB 951 GCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1010  
QY 426 TCAGCGGATCGGCTCGATGACGAGCGGCTGC 456  
DB 1011 CCCCCG 1041

RESULT 39

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

CK161715 1094 bp mRNA linear EST 05-DEC-2003  
FGAS014288 Triticum aestivum FGAS: Library 4 Gate 8 Triticum  
aestivum cDNA, mRNA sequence.  
CK161715  
CK161715.1 GI:38990184  
EST.  
Triticum aestivum (bread wheat)  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 1094)  
Allard, P., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,

Penniket, C., Roach, J.L. and Sarhan, F.  
 Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
 Unpublished (2003)  
 Contact: Wm L Crosby  
 Bioinformatics  
 University of Saskatchewan, Department of Computer Science  
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 Saskatchewan, S7N 5A9, Canada  
 Tel: 306 966 1769  
 Fax: 306 966 2033  
 Email: fgas.ests@usask.ca  
 This sequence is the direct result of the Base calling software  
 Phred (default parameters). It is the raw base calls. To aid in the  
 identification of the high quality insert the software Lucy  
 (default parameters) has been run on this sequence. Lucy identified  
 the region [40,167].  
 Plate: LAB002 row: K column: 11.  
 Location/Qualifiers  
 1. 1094  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4565"  
 /clone\_lib="Triticum aestivum FGAS: Library 4 Gate 8"  
 /note="Organ: Crown and leaf; Vector: pCMV.SPKR6; Aerial  
 parts (crown and leaf) of wheat cultivar Norstar during  
 dehydration stress. 8 mRNA populations were combined  
 before constructing the library. The first four come from  
 removing plants from vermiculite (7 day old plants) and  
 incubating them at 20C on the bench without water for 1,  
 2, 3 and 4 days. The last four come from plants grown in  
 soil in a growth chamber after watering is terminated.  
 Four samplings were taken in a two week period; the first  
 after wilting was observed and the last, two weeks later,  
 consisted of live crown and stem tissue (leaf tissue was  
 yellow and dead). First strand synthesis in this library  
 was done in the presence of methylated dCTP thereby  
 protecting from internal cleavage with NotI."

# FEATURES

source

## FEATURES

source

1. 1203  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone\_lib="BACN15E10"  
 /plasmid="pBelobAC11"  
 /note="end : 17"

## ORIGIN

Query Match 10.7%; Score 55.6; DB 14; Length 1094;  
 Best Local Similarity 42.8%; Pred. No. 26;  
 Matches 202; Conservative 0; Mismatches 266; Indels 4; Gaps 2;  
 QY 49 GCATCTCGCGGGGTGACAGGTGACAGGTGAGGTGAGCAGCAGGTGATCGCGG 108  
 DB 295 GCGGNGGNGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354  
 QY 109 CTGTGCGCGGCGATGCGGAAACACCGGGCGCAGGATCACGTGACGACGCGGCGCT 168  
 DB 355 CCGCGGGCG 414  
 QY 169 CGATCTCGAGTGGCGCGAGTGTGATCTGTGCGCGGCTTCCACACGAGGGAAC 228  
 DB 415 NCGNCCCG 474  
 QY 229 GCGTGGTGGAGCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 288  
 DB 475 GNNNGNCCCG 534  
 QY 289 CTGCGGGGTGCGGAGCGGAAACCGCGACGCGGTGAGATCCCTGTGATGCGATCG 348  
 DB 535 GCG 594  
 QY 349 TGCGGTGTGCG 406  
 DB 595 CCGCGGCG 654  
 QY 407 GCGCGGGGGCGGGGCGGTTCAGCCGATCCGATCGATGACAGCGG--CTGCGGGGT 464  
 DB 655 CGCGCGCGGGGGCG 714  
 QY 465 GCGGCTCGGCGAGATCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516

DB 715 CCG 766

RESULT 40  
 CNS015Y4/c  
 LOCUS  
 DEFINITION  
 Drosophila melanogaster genome survey sequence T7 end of BAC  
 BACN15E10 of DrosBAC library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 AL106054  
 AL106054.1 GI:5619805  
 GSS.  
 Drosophila melanogaster (fruit fly)  
 ORGANISM  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 1203)  
 Genoscope.  
 Direct Submission  
 TITLE  
 JOURNAL  
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT  
 Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the European Drosophila Genome Project (EDGP) -  
 http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
 d'Etude du Polymorphisme Humain) with funding provided by a MRC  
 project grant. The DNA was prepared from embryos by Alain Bucheton  
 and Genavieve Payan. It has been constructed in the vector  
 pBelobAC11

Location/Qualifiers  
 1. 1203  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone\_lib="BACN15E10"  
 /plasmid="pBelobAC11"  
 /note="end : 17"

ORIGIN  
 Query Match 10.7%; Score 55.6; DB 29; Length 1203;  
 Best Local Similarity 35.3%; Pred. No. 27;  
 Matches 164; Conservative 81; Mismatches 215; Indels 5; Gaps 2;  
 QY 55 CCGCGGGGTGACAGGTGACAGGTGAGGTGAGCAGCAGGTGATCGCGCGCTGCG 114  
 DB 1175 CCGCGGGCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1116  
 QY 115 CCGCATGCGGAAACACCGGGCGCAGGATCACGTGACGACGCGCGCTCGATCT 174  
 DB 1115 CCGGGGGGCGCGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1056  
 QY 175 CGACGTTGCGCGCGAGTGTGATCTGTGCGCGCGGTCCACACGAGGGGAAATGCG 234  
 DB 1055 CCG 996  
 QY 235 GTGCGACG 291  
 DB 995 CBCCCCCG 936  
 QY 292 GCGGGTGG--GAGCGGAAACGCGCGAGTGTGATGATCCCTGTGATGCGATCG 349  
 DB 935 CCBBCCG 876  
 QY 350 GCGGTGTGCG 409  
 DB 875 SGGNCCCG 816  
 QY 410 CCGGGGGCGGGGCGGGTTCAGCCGATCCGATCGATGACAGCGGTGCGGGGTG 469  
 DB 815 CCG 756





```
Db 158 GCGCAGCAGCGCGCGCGGAGACCGTATCGTACCATGCGGATGAGGGCTACGACCG 217
Qy 62 GGTGACAGGTCGACAGGTCGGAGTTGAGCAGACCGAGCTCATCGCGCTCGCGCGCAT 121
Db 218 CGAGTTCGCGAGTTGCGGACGACCTCGGCTTCAGACGGCCACCTTCCTCGGNNCCT 277
Qy 122 CCGGAAACACCGGCGCAGGATCAGGTGCGAGCAGCGCGATGCGCGCTCGATCTCGAC--G 179
Db 278 CTCNNNTGCGGACTCCATTCTGTGTGTCCTCCCGCGAGCGCGNCCNNNGCNCNN 337
Qy 180 GTGCGCGCAGCTGATCTGTGCGCGGCTCCACACCGAGGGAAGTGGTGGTGGC 239
Db 338 GCGNCCCGCGCCNCCNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 397
Qy 240 AGCGCAGCGCGCGGAGAGTCTGTCAGGCTGCGTCAAGGCTGCGTCAAGGCTCTCGCGGTCG 299
Db 398 GCGGCGCGCGCGCGCGCGCGCTTCNCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 457
Qy 300 GGAGCGGAAACGCGCAGCTGATCTGTGTCAGTGTGTCAGTGTGTCAGTGTGTC 359
Db 458 GCGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 517
Qy 360 CCGCTTGGCTGGGAGGATAGCGGTTCACGACGAGCGGACACGCGCGCGCGCGCGCG 419
Db 518 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 568
Qy 420 GCGGCTTCAAGCGATTCGCTGATGACGAGCGGCTGCGGCGTTCGCGGCGTTCGCGGAGAT 479
Db 569 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 628
Qy 480 CCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 508
Db 629 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657
```

```
RESULT 43
AG077581
LOCUS      AG077581      1086 bp      DNA      linear      GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-072C03.F, genomic survey sequence.
ACCESSION AG077581
VERSION    AG077581.1  GI:16629383
KEYWORDS   GSS.
SOURCE     Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE  1
            Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
            BAC end sequences of Library PTB
            Unpublished
REFERENCE  2
            (bases 1 to 1086)
            Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
            Direct Submission
            Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsukuba, Ibaraki, Japan
            (E-mail: chimpes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
            Tel: 81-45-503-9111, Fax: 81-45-503-9170)
            Clones are derived from the chimpanzee BAC library PTB. This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
PRIMERS
```

```
Sequencing: -21M13
LIBRARY    Vector      : PKS145
            R.Site 1   : SacI
            R.Site 2   : SacI
            Location/Qualifiers
            1. .1086
            /organism="Pan troglodytes"
            /mol_type="genomic DNA"
FEATURES
            source
COMMENT
            Clones are derived from the chimpanzee BAC library PTB. This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
PRIMERS
```

```
Sequencing: -21M13
LIBRARY    Vector      : PKS145
            R.Site 1   : SacI
            R.Site 2   : SacI
            Location/Qualifiers
            1. .1086
            /organism="Pan troglodytes"
            /mol_type="genomic DNA"
FEATURES
            source
COMMENT
```

## ORIGIN

```
Query Match      10.7%; Score 55.4; DB 29; Length 1086;
Best Local Similarity 37.7%; Pred. No. 28;
Matches 193; Conservative 0; Mismatches 315; Indels 4; Gaps 1;

Qy 1  GGTGACGACGACGACGAGTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Db 491 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 550
Qy 61  GGTGACGACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 120
Db 551 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 610
Qy 121 TGGCGGAAACACCGGCGCGAGATCACTGTCAGACGCGGATGCGCGCTCGATCTCGACGG 180
Db 611 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 670
Qy 181 TGGCGCGAGCTGATCTGCTGCGCGCGCTCCACACAGAGGGAACACTGCTCGGTGCGCA 240
Db 671 GCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 730
Qy 241 GCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 731 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 790
Qy 301 GAGCGGAAACGCGCGCGCGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Db 791 CGCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 850
Qy 361 CCTTGGCTGGAGGATAGCGGTTCAACAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 851 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 906
Qy 421 GCGGTTGAGCGGATCGCTGATGACAGCGGCTGCGGCGTTCGCGGCGTTCGCGGAGATC 480
Db 907 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966
Qy 481 CGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 512
Db 967 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 998
```

```
RESULT 44
AG072425/c
LOCUS      AG072425      1232 bp      DNA      linear      GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-063O17.F, genomic survey sequence.
ACCESSION AG072425
VERSION    AG072425.1  GI:16624227
KEYWORDS   GSS.
SOURCE     Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE  1
            Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
            BAC end sequences of Library PTB
            Unpublished
REFERENCE  2
            (bases 1 to 1232)
            Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
            Direct Submission
            Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsukuba, Ibaraki, Japan
            (E-mail: chimpes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
            Tel: 81-45-503-9111, Fax: 81-45-503-9170)
            Clones are derived from the chimpanzee BAC library PTB. This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
PRIMERS
```





Plate: LCM2004 row: b column: 13  
 High quality sequence stop: 206.  
 Location/Qualifiers  
 1. 1273  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NIH MGC 41"  
 /tissue="amelanotic melanoma, cell line"  
 /lab\_host="DHI08 (phage-resistant)"  
 /notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."

ORIGIN

Query Match 10.6%; Score 55.2; DB 12; Length 1273;  
 Best Local Similarity 50.0%; Pred. No. 31;  
 Matches 155; Conservative 0; Mismatches 154; Indels 1; Gaps 1;

202 CGCCCGGCTCCACACAGGGGAACTGGCTCGGTGGACGCGACGCCACCGGGACA 261  
 |||||  
 971 CACCCCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 912  
 |||||

262 GCTGTCAGGCTGGCTCAGGACCTCTCGCGGTGGGAGCGGAAACGCGACGGCTC 321  
 |||||

911 GCGGGGCG 852  
 |||||

322 AGATCCCTGTGTCAGTCGATCGGTCAGTGGCGGTGTCCTTCCTTGGCTGGAGATAGC 381  
 |||||

851 GCGGCG 792  
 |||||

382 GGTTCAGACAGAGCGGACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 441  
 |||||

791 CGGCG 733  
 |||||

442 ATGACACGACGCGCTCGGCGGTGGGCGGTGGGCGAGATCCGTTACCGCGCGACCGCTCG 501  
 |||||

732 CCCCCCG 673  
 |||||

502 GCGAGCGCGG 511  
 |||||

672 GGGCGCGCGG 663  
 |||||

BU400981 734 bp mRNA linear EST 27-NOV-2002  
 603481269P1 CSEQCHN59 Gallus gallus cDNA clone CHEST368j17 5', mRNA  
 sequence.  
 BU400981  
 BU400981.1 GI:25770037  
 EST.  
 Gallus gallus (chicken)  
 Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 734)  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 A Comprehensive Collection of Chicken cDNAs  
 Curr. Biol. 12 (22), 1965-1969 (2002)  
 22335534  
 12445392  
 Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology

(UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612088930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.  
 Location/Qualifiers

## FEATURES

source  
 1. 734  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone\_lib="CHEST368j17"  
 /dev\_stage="36"  
 /lab\_host="DHI08"  
 /clone\_lib="CSEQCHN59"  
 /notes="Organ: limbs; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesized C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaudo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

## ORIGIN

Query Match 10.6%; Score 55; DB 13; Length 734;  
 Best Local Similarity 49.5%; Pred. No. 31;  
 Matches 142; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 32 GCGCGCCATCATCGAACTCCCACTCCGCGGGGTGACCAAGTTCAGCCAGGTGGAGTTGAG 91  
 |||||

DB 377 GCGGAGCAGCGCGCACAGCGAGCCAGCGACAGAGCAGCGAACTTTCGCGGCGCGAG 318  
 |||||

QY 92 CAGCCAGCTCATCGCGGCTGCGCGGATCGGAAACACCGCGCCAGGATCAGTCGAG 151  
 |||||

DB 317 CAGAGCAGCGCGCGGTACAGCGCGCCAGCCGAGCAGCGCGCGCGCGCGCGCGCG 258  
 |||||

QY 152 CACGCGCGATGCGCGCTCGATCTCGACGTCGCGCGCGAGTCGATCTCGTCGCGCGGCTC 211  
 |||||

DB 257 CAGCG 198  
 |||||

QY 212 CCACACAGCGGGAACCTGGCTCGGTGGAGCGGAGCGCCAGCGCGGACAGCTGCTCAG 271  
 |||||

DB 197 GCGCGCCACGCGCCAGCGCGCGCTCGGAGGCGCTCGCGCCAGCGTCCGCGCC 138  
 |||||

QY 272 GCTGCGCTCAGCGACTCTCGCGGTCGCGAGCGGAAACGCGCACGG 318  
 |||||

DB 137 GCGCTCGCGGCTCGTCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGG 91  
 |||||

## RESULT 50

AG058001/c  
 LOCUS  
 DEFINITION Pan troglodytes DNA, clone: PTB-044M04.F, genomic survey sequence.  
 AG058001  
 ACCESSION  
 VERSION AG058001.1 GI:16595462  
 KEYWORDS GSS.  
 SOURCE Pan troglodytes (chimpanzee)  
 ORGANISM Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 1  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 BAC end sequences of Library PTB  
 Unpublished  
 REFERENCE  
 2 (bases 1 to 1285)

**AUTHORS** Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimbesc@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
**COMMENT** Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.  
**PRIMERS**  
Sequencing: -21M13  
**LIBRARY**  
Vector : PKS145  
R.Site 1 : SacI  
R.Site 2 : SacI.  
Location/Qualifiers  
1.1285  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="PTB-044M04.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
**ORIGIN**  
Query Match 10.6%; Score 55; DB 29; Length 1285;  
Best Local Similarity 46.7%; Pred. No. 34;  
Matches 195; Conservative 0; Mismatches 219; Indels 4; Gaps 1;  
QY 102 ATCGCGGCTGCGCGGATCGGAAACACCGGCGCAGGATCAGTGCACGACGCGGATG 161  
Db |||||  
553 ACCGCGCGCGGCG 494  
QY 162 CGCGCTCTGATCTCGACGCTCGCGCGCAGCTCGATCTCGTCCCGCGCTCCACACCGAG 221  
Db |||||  
493 CG 434  
QY 222 GGAAGTGGCTCGTGGGAGCGGCGAGCCCGACGCGGACAGCTCGTCCAGGCTGGCGTCA 281  
Db |||||  
433 CCGNCG 374  
QY 282 CGGACTCTCTCGCGGCTCGGAGCGGAAACCGCGCACGCTCAGATCCCTGTCAGTCCGATC 341  
Db |||||  
373 CCCCTCG 318  
QY 342 GGCCTAGTCCGCTCGTCCCTTTGGCTTGGCTTGGAGAGATAGCGGTTACAGAGCGGACC 401  
Db |||||  
317 GCCCG 258  
QY 402 ACGGCGGCGGCGGCGGCGGCGGTTTCAGCGGATCCGCTTCGATGACACGCGGCTCGGGGT 461  
Db |||||  
257 GCGCGCGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 198  
QY 462 CGGCGGCGGTTCGCGGAGATCCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 519  
Db |||||  
197 CGGCG 140

Search completed: June 27, 2004, 20:50:49  
Job time : 2115.57 secs



Micromonosporineae; Micromonosporaceae; Micromonospora.  
 1  
 Farnet, C.M., Zazopoulos, E. and Staffa, A.  
 Compositions and methods for identifying and distinguishing  
 orthosomycin biosynthetic loci  
 Patent: WO 02079505-A 229-10-OCT-2002  
 Ecopia Biosciences Inc. (CA)  
 Location/Qualifiers  
 1..37116  
 /organism="Micromonospora carbonacea"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:47853"  
 /note="var. africana"  
 ORIGIN  
 Query Match 100.0%; Score 500; DB 6; Length 37116;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-56; Indels 0; Gaps 0;  
 Matches 500; Conservative 0; Mismatches 0;  
 QY 1 GGTACCCGACCGTGTCCCGAACAACAGATCGAGATAGCGGAGAGAACACCCCGGT 60  
 Db 30486 GGTACCCGACCGTGTCCCGAACAACAGATCGAGATAGCGGAGAGAACACCCCGGT 30427  
 QY 61 AGTCCGGGTAGAGCGGTGGCGCGAAGCGTACGCGCTTCGACGCTCAGCGGCGGCGG 120  
 Db 30426 AGTCCGGGTAGAGCGGTGGCGCGAAGCGTACGCGCTTCGACGCTCAGCGGCGGCGG 30367  
 QY 121 ACACCGCGGAGTCACTGCTGCTACGTACGCGGCGGAGAGTACAGATCCATGTCGCGCA 180  
 Db 30366 ACACCGCGGAGTCACTGCTGCTACGTACGCGGCGGAGAGTACAGATCCATGTCGCGCA 30307  
 QY 181 GCCCGGCGGAACCTCCTGCTCCTTCGCCCATGATCTCGTCCGCGGTTCGACGCGAAGAGC 240  
 Db 30306 GCCCGGCGGAACCTCCTGCTCCTTCGCCCATGATCTCGTCCGCGGTTCGACGCGAAGAGC 30247  
 QY 241 AGCGGTAGTCCACCGCTCGGCGGTGAACCGTCCGCGGTTCGACGCGGAGATGCGTG 300  
 Db 30246 AGCGGTAGTCCACCGCTCGGCGGTGAACCGTCCGCGGTTCGACGCGGAGATGCGTG 30187  
 QY 301 CCGGGGTGAGCGCGGCGCTGCTGGCGGCGGTGCTGTCGACACCCAGAGACACAGGTCC 360  
 Db 30186 CCGGGGTGAGCGCGGCGCTGCTGGCGGCGGTGCTGTCGACACCCAGAGACACAGGTCC 30127  
 QY 361 GGACCGATGCCGAGAGTTCGTACGCTGGCGGTTCGCGCTCGCGCGGTACGCCACC 420  
 Db 30126 GGACCGATGCCGAGAGTTCGTACGCTGGCGGTTCGCGCTCGCGCGGTACGCCACC 30067  
 QY 421 ACCCGTTGCCCTCGGCGCTTACGAGTTACAGCGGCGAGCAGTTCGTCGGATGCC 480  
 Db 30066 ACCCGTTGCCCTCGGCGCTTACGAGTTACAGCGGCGAGCAGTTCGTCGGATGCC 30007  
 QY 481 TCGACGTCCGCGCGAACCT 500  
 Db 30006 TCGACGTCCGCGCGAACCT 29987  
 RESULT 2  
 AX195929  
 LOCUS 109519 bp DNA linear PAT 28-AUG-2001  
 DEFINITION Sequence 1 from Patent WO0151639.  
 AX195929  
 ACCESSION AX195929  
 VERSION AX195929.1 GI:15386161  
 KEYWORDS  
 SOURCE Micromonospora carbonacea  
 ORGANISM Micromonospora carbonacea  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Micromonosporineae; Micromonosporaceae; Micromonospora.  
 1  
 Hosted, T.J., Horan, A.C. and Wang, T.X.  
 Evernimycin biosynthetic genes  
 Patent: WO 0151639-A 119-JUL-2001;  
 Schering Corporation (US)  
 Location/Qualifiers  
 1..1251  
 /organism="Micromonospora carbonacea"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:47853"  
 /note="var. africana"  
 ORIGIN  
 Query Match 73.8%; Score 369; DB 6; Length 1251;

1..109519  
 /organism="Micromonospora carbonacea"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:47853"  
 ORIGIN  
 Query Match 100.0%; Score 500; DB 6; Length 109519;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-56; Indels 0; Gaps 0;  
 Matches 500; Conservative 0; Mismatches 0;  
 QY 1 GGTACCCGACCGTGTCCCGAACAACAGATCGAGATAGCGGAGAGAACACCCCGGT 60  
 Db 1 GGTACCCGACCGTGTCCCGAACAACAGATCGAGATAGCGGAGAGAACACCCCGGT 60  
 QY 61 AGTCCGGGTAGAGCGGTGGCGCGAAGCGTACGCGCTTCGACGCTCAGCGGCGGCGG 120  
 Db 61 AGTCCGGGTAGAGCGGTGGCGCGAAGCGTACGCGCTTCGACGCTCAGCGGCGGCGG 120  
 QY 121 ACACCGCGGAGTCACTGCTGCTACGTACGCGGCGGAGAGTACAGATCCATGTCGCGCA 180  
 Db 121 ACACCGCGGAGTCACTGCTGCTACGTACGCGGCGGAGAGTACAGATCCATGTCGCGCA 180  
 QY 181 GCCCGGCGGAACCTCCTGCTCCTTCGCCCATGATCTCGTCCGCGGTTCGACGCGAAGAGC 240  
 Db 181 GCCCGGCGGAACCTCCTGCTCCTTCGCCCATGATCTCGTCCGCGGTTCGACGCGAAGAGC 240  
 QY 241 AGCGGTAGTCCACCGCTCGGCGGTGAACCGTCCGCGGTTCGACGCGGAGATGCGTG 300  
 Db 241 AGCGGTAGTCCACCGCTCGGCGGTGAACCGTCCGCGGTTCGACGCGGAGATGCGTG 300  
 QY 301 CCGGGGTGAGCGCGGCGCTGCTGGCGGCGGTGCTGTCGACACCCAGAGACACAGGTCC 360  
 Db 301 CCGGGGTGAGCGCGGCGCTGCTGGCGGCGGTGCTGTCGACACCCAGAGACACAGGTCC 360  
 QY 361 GGACCGATGCCGAGAGTTCGTACGCTGGCGGTTCGCGCTCGCGCGGTACGCCACC 420  
 Db 361 GGACCGATGCCGAGAGTTCGTACGCTGGCGGTTCGCGCTCGCGCGGTACGCCACC 420  
 QY 421 ACCCGTTGCCCTCGGCGCTTACGAGTTACAGCGGCGAGCAGTTCGTCGGATGCC 480  
 Db 421 ACCCGTTGCCCTCGGCGCTTACGAGTTACAGCGGCGAGCAGTTCGTCGGATGCC 480  
 QY 481 TCGACGTCCGCGCGAACCT 500  
 Db 481 TCGACGTCCGCGCGAACCT 500  
 RESULT 3  
 AX574152/c  
 LOCUS 1251 bp DNA linear PAT 07-JAN-2003  
 DEFINITION Sequence 232 from Patent WO02079505.  
 AX574152  
 ACCESSION AX574152  
 VERSION AX574152.1 GI:27551653  
 KEYWORDS  
 SOURCE Micromonospora carbonacea  
 ORGANISM Micromonospora carbonacea  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Micromonosporineae; Micromonosporaceae; Micromonospora.  
 1  
 Farnet, C.M., Zazopoulos, E. and Staffa, A.  
 Compositions and methods for identifying and distinguishing  
 orthosomycin biosynthetic loci  
 Patent: WO 02079505-A 232-10-OCT-2002;  
 Ecopia Biosciences Inc. (CA)  
 Location/Qualifiers  
 1..1251  
 /organism="Micromonospora carbonacea"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:47853"  
 /note="var. africana"  
 ORIGIN  
 Query Match 73.8%; Score 369; DB 6; Length 1251;



Best Local Similarity 100.0%; Pred. No. 5.2e-39; Mismatches 0; Indels 0; Gaps 0;

Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 TCAGCTCGTCACGTGTACGCGGGGACGTACAGATCCACTGTCCGCCACCGCGCGGAA 191

DB 1251 TCAGCTCGTCACGTGTACGCGGGGACGTACAGATCCACTGTCCGCCACCGCGCGGAA 1192

QY 192 CTCCTGCTTCCTCCCATGATCTCGTCGGGTGTTCCAGGCGAAGAGCAGCGGTAGTC 251

DB 1191 CTCCTGCTTCCTCCCATGATCTCGTCGGGTGTTCCAGGCGAAGAGCAGCGGTAGTC 1132

QY 252 CACGCGGTTCGGGCGGTGAACCGCTCCCGGGGTGCGCACCGGATGTGCGTCCGCGGGGTGAG 311

DB 1131 CACGCGGTTCGGGCGGTGAACCGCTCCCGGGGTGCGCACCGGATGTGCGTCCGCGGGGTGAG 1072

QY 312 CGGCGCTTCCTGGCGCGGTGCTGTGTCGACACCCAGAGACAGGTCCGCGCGGATGCC 371

DB 1071 CGGCGCTTCCTGGCGCGGTGCTGTGTCGACACCCAGAGACAGGTCCGCGCGGATGCC 1012

QY 372 GCAGAGTTCGTACGCGTGGCGTCTTCGCGCGTGGCGCGTACGCCACACCGCGTTGCC 431

DB 1011 GCAGAGTTCGTACGCGTGGCGTCTTCGCGCGTGGCGCGTACGCCACACCGCGTTGCC 952

QY 432 CTCGCGCTTCAGCGAGTTCAGCGAGGCGAGCGAGTTCGCGGATGCCCTCGAGTCGCG 491

DB 951 CTCGCGCTTCAGCGAGTTCAGCGAGGCGAGCGAGTTCGCGGATGCCCTCGAGTCGCG 892

QY 492 GCGGAACCT 500

DB 891 GCGGAACCT 883

RESULT 4  
AX195930/c  
LOCUS AX195930 1248 bp DNA linear PAT 28-AUG-2001

DEFINITION Sequence 2 from Patent WO0151639.

ACCESSION AX195930

VERSION AX195930.1 GI:15386162

KEYWORDS Micromonospora carbonacea

SOURCE Micromonospora carbonacea

ORGANISM Micromonospora carbonacea

REFERENCE 1

AUTHORS Hosted, T.J., Horan, A.C. and Wang, T.X.

TITLE Evernimycin biosynthetic genes

JOURNAL Patent: WO 0151639-A 2 19-JUL-2001;

SCHERING Corporation (US)

FEATURES

source

1..1248

/organism="Micromonospora carbonacea"

/mol\_type="unassigned DNA"

/db\_xref="taxon:47853"

CDS

1..1248

/note="unnamed protein product; evdA"

/codon\_start=1

/transl\_table=1

/protein\_id="CAC59991.1"

/db\_xref="GI:15386163"

/translation="MTSDAGTTPRGTCACGGTTAQFDLGRPLSDRLFTDQLAD  
EFFFLAVGACGCTVMOLMEVEPREEMFHODYPHSSGSSVMOKHPADTARHLLEH  
ATPDPPVEIGNDVGMRLTVREAGVRLHGFSPGGVADVARGHGVTVTFPEEST  
ARRVETGPDADVIFRANTICHIPLYDILGDLALLAPDGVFVFPDLYLDLKT  
PDQYDEHFFFSARSQVLAARFGPALVDVERLPVHGGVRYVTIARAGARQPAERVA  
ALIGHEDARGVADSRDLRFADVEGIRLTDLALLSLKAEKGVWYAGTAKSATVT  
NFCGGPDLVSWCDRTTPAKQGRLTTPGTHIPVTPDPAVDVALLFAWNHAEIM  
AKEQEFRRAGQWILYVPRVHTS"

## ORIGIN

Query Match 73.2%; Score 366; DB 6; Length 1248;

Best Local Similarity 100.0%; Pred. No. 1.3e-38; Mismatches 0; Indels 0; Gaps 0;

Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 GCTCGTCACGTGTACGCGGGGACGTACAGATCCACTGTCCGCCACCGCGCGGAACTC 194

DB 1248 GCTCGTCACGTGTACGCGGGGACGTACAGATCCACTGTCCGCCACCGCGCGGAACTC 1189

QY 195 CTGCTCCTTCCTCCCATGATCTCGTCGGGTGTTCCAGGCGAAGAGCAGCGGTAGTCAC 254

DB 1188 CTGCTCCTTCCTCCCATGATCTCGTCGGGTGTTCCAGGCGAAGAGCAGCGGTAGTCAC 1129

QY 255 CGGCTCGGGCGTGAACCGGTCCGGGTGCGCACCGGATGTGCGTCCGCGGGGTGAGCGG 314

DB 1128 CGGCTCGGGCGTGAACCGGTCCGGGTGCGCACCGGATGTGCGTCCGCGGGGTGAGCGG 1069

QY 315 GCCCTGCTTCGCGCGGTGCTGTGTCGACACCCAGAGACACAGGTCCGAGTCCGCGCA 374

DB 1068 GCCCTGCTTCGCGCGGTGCTGTGTCGACACCCAGAGACACAGGTCCGAGTCCGCGCA 1009

QY 375 GAAGTTCGTCAGGTGGCGTCTTCGCGCGTGGCGGTACGCCACACCGCTTTCGCGTC 434

DB 1008 GAAGTTCGTCAGGTGGCGTCTTCGCGCGTGGCGGTACGCCACACCGCTTTCGCGTC 949

QY 435 GGCCTTGAGCGAGTTCAGCAGGCGGAGCAGGTTCGCGGATGCCCTCGAGTCCGCGGC 494

DB 948 GGCCTTGAGCGAGTTCAGCAGGCGGAGCAGGTTCGCGGATGCCCTCGAGTCCGCGGC 889

QY 495 GAACT 500

DB 888 GAACT 883

RESULT 5  
AX574202/c

LOCUS AX574202 10035 bp DNA linear PAT 07-JAN-2003

DEFINITION Sequence 282 from Patent WO02079505.

ACCESSION AX574202

VERSION AX574202.1 GI:27551681

KEYWORDS Micromonospora carbonacea

SOURCE Micromonospora carbonacea

ORGANISM Micromonospora carbonacea

REFERENCE 1

AUTHORS Farnet, C.M., Zazopoulos, E. and Staffa, A.

TITLE Compositions and methods for identifying and distinguishing

JOURNAL orthosomycin biosynthetic loci

Patent: WO 02079505-A 282 10-OCT-2002;

Ecopia Biosciences Inc. (CA)

FEATURES

source

1..10035

/organism="Micromonospora carbonacea"

/mol\_type="unassigned DNA"

/db\_xref="taxon:47853"

ORIGIN

Query Match 59.7%; Score 298.4; DB 6; Length 10035;

Best Local Similarity 76.1%; Pred. No. 4.1e-30;

Matches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;

QY 1 GGTACCCGACCGTGTCCCGGAACAACAGAGTCGAGATACGGCGAGAGGAAACACCCCGGT 60

DB 3742 GATATCCGAGCTCTCGTGAAGACCGAGTCCAGGTACGGGACACGAGACACCCCGG 3683

QY 61 AGTCCGGGTAGCGGTGGCGCGAAGCGGTACCGCGCTTCGACGTCACGGCGCGGCGG 120

DB 3682 GGTCCGGAAGACCCCGGGGTGAAGACGTACCGCGCTTCGACGCGCGAGTTCGTTGCTA 3623

QY 121 ACACCC---GGGGGTTCAGTTCGACGTGTACGGGGGAGCGTACAGGATCCACTGTCC 176

DB 3622 CTTGCACGGCCCCACTCAATCCCGACGTGAACGTGCGGAAAGTACAGGATCCAGCCCC 3563

QY 177 GCCAGCCCCCGGAACTCCTGTCTTCCCATGATCTCGTGGGCGGTTCACGCGGAA 236

DB 3562 GCCGGCTTCGCGGAACGCTGTCTCTTCGCCATGATCTCGTGGGCGGTTCACGCGGAA 3503

QY 237 GAGCAGCGCTAGTCCACCGGTCGGGCTGAAACGGCTCCGGGTGCGCAACCGGGATGTG 296  
 Db 3502 GAGCAGCGCTAGTCCCGGCGGTGGTCGGGAATCTGTCGCGCATGAACCGGGATGTG 3443  
 QY 297 CGTCCGGGGGTGAGCGCGCCCTGTTGGCCGGGCTGCTGTCGACACACCGAGGACAG 356  
 Db 3442 CGTCCCGGGGTGAGCGCGCCCTGTTGGCCGGGCTGCTGTCGACACACCGAGACAG 3383  
 QY 357 GTCGGGACCGATGCGGACGAGAGTTCGTACCGTGGCGCTCTTCGCGCTCGCGCGGTACGC 416  
 Db 3382 GTCGGGACCGATGCGGACGAGAGTTCGTACCGTGGCGCTCTTCGCGCTCGCGCGGTAGGC 3323  
 QY 417 CACACCGCGCTTGGCTCGGCTTGGAGGAGTTCAGAGGGCGAGAGTTCGGTTCGGAT 476  
 Db 3322 CACACCGCGCTTGGCTCGGCTTGGAGGAGTTCAGAGGGCGAGAGTTCGGTTCGGAT 3263  
 QY 477 GCGCTCGACGTCGGGGCGAAC 498  
 Db 3262 CCGGCGGACCTGGGCGAGCGAAC 3241

RESULT 6  
 AX205014/c  
 LOCUS AX205014 11115 bp DNA linear PAT 30-AUG-2001  
 DEFINITION Sequence 49 from Patent WO0155180.  
 ACCESSION AX205014  
 VERSION AX205014.1 GI:15394258  
 KEYWORDS  
 ORGANISM  
 Micromonospora carbonacea  
 Micromonospora carbonacea  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Micromonosporineae; Micromonosporaceae; Micromonospora.



REFERENCE  
 AUTHORS Staffa, A., Zazopoulos, E., Mercure, S. and Nowacki, P.  
 TITLE Genetic locus for evernimycin biosynthesis  
 JOURNAL Patent: WO 02079505-A 230 10-OCT-2002  
 Ecopia Biosciences Inc. (CA); Farnet, Chris (CA)  
 FEATURES  
 Location/Qualifiers  
 1..11115  
 /organism="Micromonospora carbonacea"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:47853"

misc\_feature  
 /note="ORF 41 (positive strandedness) incomplete;  
 C-terminus only"  
 1213..2331  
 /note="ORF 42 (positive strandedness)"  
 2364..3611  
 /note="ORF 43 (positive strandedness)"  
 3623..4243  
 /note="ORF 44 (positive strandedness)"  
 4149..5177  
 /note="ORF 45 (positive strandedness)"  
 5177..6094  
 /note="ORF 46 (negative strandedness)"  
 6271..7824  
 /note="ORF 47 (negative strandedness)"  
 7903..8760  
 /note="ORF 48 (negative strandedness)"  
 8781..9800  
 /note="ORF 49 (negative strandedness)"  
 ORIGIN

Query Match 59.7%; Score 298.4; DB 6; Length 11115;  
 Best Local Similarity 76.1%; Pred. No. 4e-30;  
 Matches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;  
 QY 1 GGTACCGGCTGTTCGGGACACAGAGTTCGAGATCGGCGAGAGGAGAACCCCGCGT 60  
 Db 3749 GATATCCGAGCGTCTCGGTGAGACCGAGTCCAGGTACGGGACACGAGAACCCCGCG 3690  
 QY 61 ACTCCGGGTAGACGGTGGCGGCGAAGCGGTACGCGCTTCGACGCTGACGGCGGGCGG 120

Db 3689 GGTCCGGGAAGACCCCGGGGTGAAGAGCGTACGCGCCCTCGACGGCGAGTTCCGTTGCTA 3630  
 QY 121 ACACC-----GGCGGGTACGTCGTCACGTGTACGCGGGGACGTACAGATCCACTGTCC 176  
 Db 3629 CTGTACGCGCCCACTCAATCCGCACTGTGAACGTGCGGAACGTACAGATCCAGGCC 3570  
 QY 177 GCGAGCCCGCGGAACTCTCTGCTCTTCCCATGATCTGTCGGCGTGTTCAGCGCGAA 236  
 Db 3569 GCGGCGCTTCCCGGAACGCGCTCTCTTCCCATGATCTGTCGGCGTGTTCAGCGCGAA 3510  
 QY 237 GAGCAGCGGTAGTCCACCGGTCGGGGTGAACCGTCCGGGGTGGCGACCGGATGTG 296  
 Db 3509 GAGCAGCGGTAGTTCGGGGTTCGTGCGGAACTCTGTCGCGCATGAACCGGATGTG 3450  
 QY 297 CBTGCGGGGGTGAAGCGGCGCTCTGTCGGCGCTGTCGTCACACCGAGACAG 356  
 Db 3449 CBTGCGGGGGTGAAGCGGCGCTCTGTCGGCGCTGTCGTCACACCGAGACAG 3390  
 QY 357 GTCCGAGCCGATGCGCGAGAGTTCGTACGCGGGCTCTTCGCGCTCGCGCGTACGC 416  
 Db 3389 GTCCGCGCGATGCGCGAGAGTTCGCCACGCTCGGCTCTTGGCGTCCGCCGTAGGC 3330  
 QY 417 CACCAACCGCTTCCCTCGGCTTGAAGCGAGTTCAGCAGGGCGAGCAGTCCGTCGGAT 476  
 Db 3329 CACCAACCGCTTCCCTCGGCTTGAAGCGAGTTCAGCAGGGCGAGCAGTCCGTCGGAT 3270  
 QY 477 GCGCTCGACGTCGGGGCGAAC 498  
 Db 3269 CCGGCGGACCTGGGCGAGCGAAC 3248

RESULT 7  
 AX574150/c  
 LOCUS AX574150 1251 bp DNA linear PAT 07-JAN-2003  
 DEFINITION Sequence 230 from Patent WO02079505.  
 ACCESSION AX574150  
 VERSION AX574150.1 GI:27551652  
 KEYWORDS  
 SOURCE Micromonospora carbonacea  
 ORGANISM  
 Micromonospora carbonacea  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE  
 AUTHORS Farnet, C.M., Zazopoulos, E. and Staffa, A.  
 TITLE Compositions and methods for identifying and distinguishing  
 orthosomycin biosynthetic loci  
 JOURNAL Patent: WO 02079505-A 230 10-OCT-2002;  
 Ecopia Biosciences Inc. (CA)  
 FEATURES  
 Location/Qualifiers  
 1..1251  
 /organism="Micromonospora carbonacea"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:47853"

ORIGIN  
 Query Match 51.3%; Score 256.6; DB 6; Length 1251;  
 Best Local Similarity 81.2%; Pred. No. 1.9e-24;  
 Matches 298; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
 QY 132 TCAGCTCGTCACGTGTACGCGGGGACGTACAGGATCCACTGTCCGCCAGCCCGCGGAA 191  
 Db 1251 TCATCCGCGACGTGAACGTGCGGAACGTACAGGATCCAGGCCCGCGCTCCGGA 1192

QY 192 CTCTGCTCTTCGCCATGATCTGTCGGCGGTTCAGGCGAGAGCAGCGGTAGTC 251  
 Db 1191 CGCTGCTCTTCGCCATGATCTGTCGGCGGTTCAGGCGAGAGCAGCGGTAGTC 1132  
 QY 252 CACCGGTGCGGGGTCAACCGTCCGGGGTGGCGCACCGGATGTGCTGCCGGGGTGA 311  
 Db 1131 CGCGGGGTGGTCCGGAATCTGCTCCCGCATGAACCGGGATGTGCTGCCGGGGTCA 1072  
 QY 312 CGGCGCTGCTTGGCGGGGTGCTGTCGCACACCGCAGGAGACCGAGTCCGACCGATGCC 371

||||| 1071 GCGGCCCTGTTGGCGGGCTGCTGCTGCTACACCGCGACACCGAGTCCGCCGATGCC 1012  
||| 372 GCAGAGTTTCGTCACGGTGGCGCTCTTCGCCGTCGCGGACGCGACACCGCTGCC 431  
||| 1011 GCAGAGTTTCGTCACCGTGGCGCTCTTCGCGTCCGCGTAGGCCACACCGCTTGC 952  
||| 432 CTCGCGCTGACGAGTTCAGACGGCGAGCGTCCGTCGCGATGCCCTGACGTCGCC 491  
||| 951 CTCGCCGCTGACCGCTTCAGACCGCGAGTCCGTCGCGATGCCCTGACGTCGCC 892  
||| 492 GCGCAAC 498  
||| 891 AGCGAAC 885  
|||

RESULT 8  
AOPCZA361/c 37941 bp DNA linear BCT 06-FEB-2003  
LOCUS AMycolatopsis orientalis cosmid PCZA361.  
DEFINITION AMycolatopsis orientalis  
ACCESSION AJ223998  
VERSION AJ223998.1 GI:2894153  
KEYWORDS  
SOURCE  
ORGANISM  
Amycolatopsis orientalis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.  
1  
van Wageningen, A., Kirkpatrick, P., Williams, D., Harris, B.,  
Kershaw, J., Leonard, N., Jones, M., Jones, S., and Solenberg, P.  
Sequencing and analysis of genes involved in the biosynthesis of a  
vancomycin group antibiotic  
*Chem. Biol.* 3: 155-162 (1998)  
2 (bases 1 to 37941)  
Kershaw, J.  
Direct Submission  
Submitted (12-FEB-1998) Sanger Centre, Hinxton, Cambridge CB10 1RQ,  
England. Contact sj@sanger.ac.uk  
Location/Qualifiers  
1 .37941  
/organism="Amycolatopsis orientalis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:31958"  
/clone="PCZA361"  
<1 .1014  
/notes="similar to p450 related oxidase/hydroxylase"  
/codon\_start=1  
/transl\_table=1  
/product="PCZA361.27"  
/protein\_id="CAA11791.1"  
/db\_xref="GI:2894184"  
/db\_xref="GOA:O52816"  
/db\_xref="SPTREMBL:O52816"  
/translations="DHEFTTRPQFTHAKSDAHVEAQFVGQI:STYDPPHTRLRKMLT  
PEFTVRIRMEPAIOGLIDRLDMVEAGPPADLQGLFADPVGALACELGIPRDD  
ORREVRIRRENTDLSRLKRAADSAAFNRYLDNLSRORQDPDEGLGMI:VREHGDN  
VTBEEKGLTALILGQVFAVAGMIGVLALENPDQIQLLAFAGPEKADRVNVLVR  
YLSFQAPNRLAKOVVIDGQIKAGDVYLCVLMANRDEALTNPVNLNANRAVS  
DVGFHGHIYCVGAALARSMLRWAYOTLMORFGLRLAVPIAEVKYRSAPVDCPQDV  
VTW"  
1076 .2551  
/notes="similar to non-heme oxygenase/halogenase"  
/codon\_start=1  
/transl\_table=1  
/product="PCZA361.26"  
/protein\_id="CAA11780.1"  
/db\_xref="GI:2894173"  
/db\_xref="GOA:P96557"  
/db\_xref="SPTREMBL:P96557"  
/translations="MSVDEDFVVGAGGPGAGSTVATLVAMQGHVRLLEKPEPPYQOI  
GESLLPATVHCVCWELGITDELANAGPPVKRGTFKMGAPPEPTWTFHGISAKMAGST  
SHAYQVBRARPEDELLNNAKKGVYREGSPVTDVDEGERTGLAYTADGNGREYS  
ARFVIDASGNKRLYSKVGSRNYSFRLALFGYFEGGRULPAPVSGNLSVAFDS

GMWYIPLSDTLTSGAVVRBDAEKIQDREKALNALIAECPLISEYLANATRVTTTG  
KYGLRVKDYISYQQTYYRPGMILIGDAACFDVDFSSGVLHATYSALLAARSINSV  
LAGDEKALNEFEMRYRREYFIEFLVSYFQMVNVSGETSPWQAKSTVVKQAWQ  
SFVELLGGVSSSETALTAAADRIAARSFAAAVDQASGDDNMFNFKSTVVKQAWQ  
EAQVQMKALLGEDAEPELPPGGLVTSFDGMMKMLPHHPA"  
2596 .3786  
/notes="similar to glycosyltransferase"  
/codon\_start=1  
/transl\_table=1  
/product="PCZA361.19"  
/protein\_id="CAA11774.1"  
/db\_xref="GI:2894167"  
/db\_xref="GOA:P96558"  
/db\_xref="SPTREMBL:P96558"  
/translations="MRVLITGCGSRGDTPLVALAARLREGADARMCLPPDYVERCA  
EVGVPVYGRVAVRAGAREGELPGCAAEVTVVAEWPDKVPAALIEGCDVVTGILL  
PAVAVRSMAEKLGIPRYTYVLPDHLPEQSQAERDMXNQADRLLFGDAVNSHRASI  
GLPVEHLYDYGTDQPLWLAADPVLSPFLRDTGTVQTKANWILPDERPLSAELEFLA  
AGSTPVYVGSSSRPATADAAMKAKAVASGRRIVLGRVWADLVLPDDGACDFVVG  
ENVLQELFGRVAAAIHDSAGTILLARAGIQIVVVRVVDNVDEYHADRVARLGV  
GVAVDGVPVTDLSLSAALDTALAPETRARATTVADTIRADGTTVAQALLFDVLSLEKP  
TVPA"  
3821 .5044  
/codon\_start=1  
/transl\_table=1  
/product="PCZA361.20"  
/protein\_id="CAA11775.1"  
/db\_xref="GI:2894168"  
/db\_xref="GOA:P96559"  
/db\_xref="SPTREMBL:P96559"  
/translations="MRVLITGCGSRGDTPLVALAVRDLGADVRMCAPDCAERLA  
EVGPHVPGVPSAPIQORAKPLTAEDVRFTTEATATQDEIPAAEGCAAVTTGCL  
LAAAGVRSVAEKLGIPYFAFHCPSYSPSYFPPLPGPSTQDTIDIPAQERNNQ  
SAYQRYGLNSHRDAIGLPPVEDIFTGYTDHPWVAADPVLAPLQPTDLDAVQTGW  
ILPDERPLSAELEFLAAGSPPVYLGFSAGSGCIDDAARVAIEAIAHGRRIVLLSG  
DLVLPDDGACDFVDEVLQVLFSAARAIHSGSAGTEHLATLGIPTQIVIPRHTD  
QPYAERVAEDIGVAGLEGGVPTFDMSAAVATAPETRARATAVAGTITDGAAVA  
ARLLDAVSRKSAVLA"  
6495 .7721  
/notes="similar to hydroxylase"  
/codon\_start=1  
/transl\_table=1  
/product="PCZA361.22"  
/protein\_id="CAA11777.1"  
/db\_xref="GI:2894170"  
/db\_xref="GOA:P96561"  
/db\_xref="SPTREMBL:P96561"  
/translations="MSVTQSCRICDGTVKEPITDFGROPLSDAFVVPDGKGEFFFLA  
TGI CDSCVTQVLMEEVPRDLMFHDAYPVLSSGSAFMRTFHELAKRLLTTLTGCDPF  
IVELGCDNGIMLKMAEAGVRLQGVPSGVSADLAAGKIVRKDPFEETAAADIRET  
DGPADVIAANTLCHI PMDSILKGVTTLLGPGVTFPFDYLGDI:VERTSPODIYE  
HFFFTFASVQEMAKRNGLELVDERIPVHGGEVRYTIALAGAKOPTFAVALLAWEA  
BRKLAETLREAFATNVKKIKDDLIATLTKLRAEGRVVGYGATAKATVNFCGITP  
DLVEFI:STTAKQNRILSPGQHI PVRSPKEFAADYPDYVALLFAWNHDEIMNAEQAFR  
DAGGQWILYVNVHVS"  
7732 .8556

/codon_start=1		Query Match	43.2%; Score 216; DB 1; Length 37941;
/transl_table=11		Best Local Similarity	75.0%; Pred. No. 1.4e-19;
/product="PCZA361.23"		Matches	270; Conservative
/protein_id="CAAL1778.1"			0; Mismatches 90; Indels 0; Gaps 0;
/db_xref="GI:2894171"			
/db_xref="SPTREMBL:O52804"			
/translation="MSQNLGAGLLAISPHLDDAVLSFGAGLARAAGAKVTVTVTF		QY	140 TCAGGTGTACGGGGGACAGTACAGGATCCACTGTCCGCGCAGCCCGGCGAACTCTCTGCT 199
AGTATPPSPAAERLIGWGLSPDASLHRRNEDIAALDHLGVYRHRFLDAIYRT		DB	716 TCAGGTGTACGGGGGACAGTACAGGATCCACTGTCCGCGCAGCCCGGCGAACTCTCTGCT 7657
LPGRLADNVPGKRLAISLSPDTPDLFAVRDDIKSVVEEDCTLTILCAAGNG		QY	200 CTTTCCGCGCATGATCTCGTCCGCGTGTTCACAGGCGAAGAGAGCGCTAGTCCACCGCGT 259
HINERTRDALVAHAKELPYRLMEDLPHAMFGAGPAELPSGFDLGTGDFGSGVTDM		DB	7656 CGCGTTCATGATCTCGTCCGCGTGTTCACAGGCGAAGAGAGCGCTAGTCCGCGGTAGT 7597
RDRKFAALRIPQMLMLGFGKDFQAQDDEHAKNSPGGGIGETWTFVVSREDS"			
8589..9431			
/notes="similar to methyl transferase"		QY	260 CGCGGTGAACCGCTCCGGGTGCGCACCGGGAGTGGTCCGCGGGGTGAGCCGGCCCT 319
/codon_start=1		DB	7596 CGCGGTGAACCGCTCCGGGTGCGCACCGGGAGTGGTCCGCGGGGTGAGCCGGCCCT 7537
/transl_table=11			
/product="PCZA361.24"		QY	320 GTTTCGCGCGGCTCGTGTTCGACACCCAGGAGACAGGTCCGAGCCGAGTCCGCGAGT 379
/protein_id="CAAL1779.1"		DB	7536 GTTTCGCGCGGCTCGTGTTCGAGATGAATCCACAGGTCCGCGGTGATGCCACAGAGT 7477
/db_xref="GI:2894172"			
/db_xref="GOA:O52805"			
/db_xref="SPTREMBL:O52805"		QY	380 TCGTACCGGTGGCGCTCTTCGCGCGGCTGAGCGAGTCCGAGCCAGCCAGCCGCTTCGCGCT 439
/translation="MSQNLGAGLLAISPHLDDAVLSFGAGLARAAGAKVTVTVTF		DB	7476 TCGTACCGGTGGCGCTCTTCGCGCGGCTGAGCGAGTCCGAGCCAGCCAGCCGCTTCGCGCT 7417
QADGTGEAREFATRTGPGVPLELAAGMGLTFFDLGWEVTALELSTVLAAR		QY	440 TGACGCGAGTTACAGCGGCGAGGAGTGGTCCGCGATGCCCTCGACGTCGCGCGGAGACC 499
KRLAARADVLRSLCTLVGGDMSAFDLKRGTVTVISGINSINELDEADRLGLVASVREH		DB	7416 GCAGTGTGGTACGACGCGGATCAGTGTCTCTTGATCTTTCACGTGTGCTGGCAAGC 7357
LEPGKGLDLSMSEAAESELERKOELEPGSGRRYVLAHVRLPAEIOEINHPADE			
TTDFVVCVTHRRLLAPQDVVRELVRSGFVIAQTFFASGAGRKDMVLVEAVMPCAT			
ADAR"			
complement (9400..10716)			
/notes="similar to aminotransferase"			
/codon_start=1			
/transl_table=11			
/product="PCZA361.25"			
/protein_id="CAAL1780.1"			
/db_xref="GI:2894183"			
/db_xref="GOA:O52815"			
/db_xref="SPTREMBL:O52815"			
/translation="METLVFMSYGLSTQSLMETLHSLTDPALISNNMLNELIDEPY			
VAISMAARPEEFVDRLIEYDAYCDHLRDKLAEAVVTRTFQYGTGKIAD			
LIALAEDENIDAAASVVTVCAQAMFLITLRLADRDLVLLAPTYVGLTGA			
LLTTPVWSTANGVDPELVQLKRAEQGKRVACVYTFNFAKNTGTSVDLPAR			
HRLVEAANGILLDENAIGLFSERLPSLSDRSNIVYIGSPAKTGMPCRGVY			
VVAQVAGGILLDQSKGLMLTVNTSPIAQVIAQVIAQVIAQVIAQVIAQVIAQ			
QRLQLTSELETLGLPEVNTPTGGFTVTVTFVVDVDELLAARHGVLFPT			
MEHFYGGDGFNQLRLSISLTPELKEGVTRLAALITARLWFR"			
10934..11764			
/notes="similar to haloperoxidase"			
/codon_start=1			
/transl_table=11			
/product="PCZA361.30"			
/protein_id="CAAL1784.1"			
/db_xref="GI:2894177"			
/db_xref="GOA:O52809"			
/db_xref="SPTREMBL:O52809"			
/translation="MLMTTGTGRLSHDHGDSPLVLLTGTGAPSSVWDLHQPVLRL			
AGFRVITMNRGIPSPDEGIDGTFIDLVADVAALTEHLGVAPCVVGTSMGVIQ			
ELALHPELDLVAALMAAGRSLSVORVLAAGEAKLIELGLTELPFGVLAARVNHNG			
PATLADDLTGDLDFASDNWGGVRAQLQALPDRIADYRAIKVPCHVIFSEHD			
LVAPPSAGRELAAIPGATHRTIPGCGHFGYLENPEAVNRELLRFLRAESAARLEETA			
"			
11761..13503			
/notes="similar to peptide synthetase"			
/codon_start=1			
/transl_table=11			
/product="PCZA361.18"			
/protein_id="CAAL1773.1"			
/db_xref="GI:2894166"			
/db_xref="GOA:O52803"			
/db_xref="SPTREMBL:O52803"			
/translation="WTGEAQFSTAPALFEATAAIVDPRAVAMGTTLTVAELNGEA			
NLALHRLVHGVGPKLVALAPRSVEFVIALVHAKAGAAVFPVDPYDRRRLML			
DTSACHLALPQDVAGAVVMSVERQASAPNLTDODRLGLPILPHPAYIYTS			
STGQPKGVLTFRGIPNADDTYRQNLVRSILAFAPSPFDAAVAFWFIWAGGC			
LVLAASDLIPGEPLARLVDRQITHTVLPSPALAPLEEGGLPFLGLTLVAGEACPA			



RIVESGVEPRLVAGRGTVDAALVERVEAGDGLAEIADRCADHTAARLDPPVAGVL	
VRVAVTDAGPGRAGRLVAAHLVVDVUSWRLIPDLOAACEAVVAGGPPALDPDVS	
FRWSRLDGEAIRIGELAVTEILLDQAQSLGELDPDRDVTSTVAGRSWTVPREHA	
GLVDEVTSAFHCGVDEAVVAVDVEGHRQPGGLDLSRLTG	
WFTDVEPLDVTGDPFAEAGDAGRLKQVKNVRAVPDGGGLGVMUYSILNAET	
GVLVLAAPKAEIGFNTLGRFSAGGERQWQITGVGGAEDQPLRHRVIDAVV	
DGPDPGEFTLTVWAGRLMDAEASLAQWMLAGLAHVAAGGPGGTSPDPLT	
ALQREVAERFAVAGLLDIPLSPLQEGLLFHAADRDGPDVAVASRTLAIDPLDVA	
RFRASVTLIDRHPALRASPFOLESCEAVQVIAQVDPDMRETDLSGLPESALAEFD	
RLAARHARERDLTKAPQLRHLVRLGDRRYRLIFTSHVIVADGWSLPLLVLDVLTAY	
EAGDGRTPAATSDYDFLANVDQKGAQAGQNRTELGLADEATHVPPGSIITPLE	
PERVAFELDDETSRLRVEFTRRHGVNTANTLFGVWALHARLAGRNDVVFGAAVAGRP	
Query Match 41.8%; Score 208.8; DB 1; Length 56917;	
Best Local Similarity 71.1%; Pred. No. 1e-18;	
Matches 276; Conservative 0; Mismatches 112; Indels 0; Gaps 0;	
QY	112 GCGCGCGGACACCGCGGCTACGTCGTCACGTGTACCGGGGAGCTACAGATCCAC 171
Db	38199 GACGCGAGGATTCGATGCGCGGTGCTACGTCACGTGTGCGCACGTAAGGATCCAC 38140
QY	172 TGTCCGCGACCGCGGGAATCTCTGCTCTTCCGCAATGATCTCGCGGCTGTTCACG 231
Db	38139 TGACCGCGGCGGTACAGAAAGGCTGTTCACGTTGCTGCGCGGTGTTCACG 38080
QY	232 GCGAAGACGACGCGGTAGTCACCGCGTGGGGGTGAAACCGTCCGGGTGCGCACCGG 291
Db	38079 GCGAACACGACGCGGTAGTCGCGGTGTTCGCGCGAATTCCTCGGACCGCGG 38020
QY	292 ATGTGCGTCCGCGGGGAGCGCGCTGTGTGCGCGGTGTGTGCGCACCGACCGAGG 351
Db	38019 ATGTGCTGTCCGCGGCTAGCTTGCTGCTGCGCGGTGTGTGCGCGGTGCTG 37960
QY	352 ACCAGGTCCGACCGATGCGCGCAGAGTTCGTACAGGTGGCGTCTTCGCGGTGCGG 411
Db	37959 ACCAGGTCCGCGGTGATGCGCGCAGAGTGTGTACCGTGGCACTTCTTGGCGGT 37900
QY	412 TACGCCACCGCTGCGCTGCGCTGTGAGCGAGTTCAGCGGGGAGCAGTGGTG 471
Db	37899 TAGCGACGACCGCTGCGCTGCGCTGTGAGCGAGTGTGTGAGCGCGATCAGTCTTC 37840
QY	472 CGGATCGCCTCGACGTGCGGCGGAACC 499
Db	37839 TTGATCTCTTTCAGCTGCGTGGCGGAAC 37812
RESULT 10	
SAR7932	
LOCUS 18977 bp DNA linear BCT 26-FEB-2002	
DEFINITION Streptomyces argillaceus mithramycin biosynthetic genes.	
ACCESSION AJ007932	
VERSION AJ007932.2 GI:8926774	
KEYWORDS acyl carrier protein; acyl CoA ligase; aromatase; cyclase; D-mycarose 2,3-dehydratase; D-mycarose 3-C-methyltransferase; D-oliose 2,3-dehydratase; D-oliose 4-ketoreductase; D-oliose 2,3-dehydratase; dtdp-glucose 4,6-dehydratase; dtdp-glucose synthase; ketoacyl synthase; ketoreductase; mtmc gene; mtmd gene; mtme gene; mtmI gene; mtmL gene; mtmO gene; mtmOI gene; mtmOII gene; mtmP gene; mtmQ gene; mtmS gene; mtmT gene; mtmXI gene; mtmXII gene; mtmXIII gene; mtmU gene; mtmV gene; mtmX gene; mtmY gene; oxygenase.	
SOURCE	
ORGANISM Streptomyces argillaceus	
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
REFERENCE	
AUTHORS 1	
TITLE Lombo, F., Blanco, G., Fernandez, E., Mendez, C. and Salas, J.A.	
Characterization of Streptomyces argillaceus genes encoding a polyketide synthase involved in the biosynthesis of the antitumor mithramycin	
JOURNAL Gene 172 (1), 87-91 (1996)	
MEDLINE 96257259	
PUBMED 8654997	

REFERENCE	2
AUTHORS	Lombo, F., Siems, K., Brana, A.F., Mendez, C., Bindseil, K. and Salas, J.A.
TITLE	Cloning and insertional inactivation of Streptomyces argillaceus genes involved in the earliest steps of biosynthesis of the sugar moieties of the antitumor polyketide mithramycin
JOURNAL	J. Bacteriol. 179 (10), 3354-3357 (1997)
MEDLINE	97294479
PUBMED	9150235
REFERENCE	3
AUTHORS	Prado, L., Lombo, F., Brana, A.F., Mendez, C., Rohr, J. and Salas, J.A.
TITLE	Analysis of two chromosomal regions adjacent to genes for a type II polyketide synthase involved in the biosynthesis of the antitumor polyketide mithramycin in Streptomyces argillaceus
JOURNAL	Mol. Gen. Genet. 261 (2), 216-225 (1999)
MEDLINE	99200478
PUBMED	10102355
REFERENCE	4
AUTHORS	Gonzalez, A., Remington, L.L., Lombo, F., Fernandez, M.J., Prado, L., Brana, A.F., Kunsel, B., Rohr, J., Mendez, C. and Salas, J.A.
TITLE	The mtmVOC genes of the mithramycin gene cluster in Streptomyces argillaceus are involved in the biosynthesis of the sugar moieties
JOURNAL	Mol. Gen. Genet. 264 (6), 827-835 (2001)
MEDLINE	21148776
PUBMED	11254130
REFERENCE	5
AUTHORS	Salas, J.A.
TITLE	Direct Submission
JOURNAL	Submitted (14-JUL-1998) Salas J.A., Biologia Funcional Universidad de Oviedo, Julian Claveria S/N, Asturias, SPAIN
REMARK	replaced by [6]
REFERENCE	6
AUTHORS	Salas, J.A.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUL-2000) Salas J.A., Biologia Funcional Universidad de Oviedo, Julian Claveria S/N, Asturias, SPAIN
COMMENT	On Jul 5, 2000 this sequence version replaced gi:3334818
FEATURES	Location/Qualifiers
source	1..18977
	/organism="Streptomyces argillaceus"
	/mol_type="genomic DNA"
	/strain="ATCC 12956"
	/db_xref="taxon:41951"
Gene	126..1193
	/gene="mtmD"
CDS	126..1193
	/gene="mtmD"
	/codon_start=1
	/evidence=experimental
	/transl_table=11
	/product="dtdp-glucose synthase"
	/protein_id="CAA07754.1"
	/db_xref="GI:3334819"
	/db_xref="GOA:O08245"
	/db_xref="SPTREMBL:O08245"
	/translation="MKALVLSGGSGTSLRLPTHTSPKOLVPVANKPLFVLEADIAQA
	GLTDVGVVGTAREIEADVAGDSKFGQVLYLPOEKPLGAHVLARDVLDGDDFV
	HTDGNFVVGGVGLVDFRADRPDARILLITFPDPTAFGLAELDGEVVRLEKPT
	HPKSNMVGIVLYFTPLVHEAIVRAIEPNSRGELEITHAIQMLIDNRHDSRVTHGYW
	KDTGNMVMLEVNRVTLVEGVFCTEGPDVAGEIVGRVIEAGARVSRVIGPVAVIG
	SGTVISGYSIGPYTSVSEGCRIEDSEIFSVILRDSRVGRVGRSGLIGRSVTVA
	PRKAGTSLVLDGHSQGQISS"
	1190..2185
gene	/gene="mtmE"
CDS	1190..2185
	/gene="mtmE"
	/codon_start=1
	/transl_table=11
	/product="dtdp-glucose 4,6-dehydratase"
	/protein_id="CAA07755.1"
	/db_xref="GI:3334820"
	/db_xref="GOA:O08246"
	/db_xref="SPTREMBL:O08246"

```

/translation="MTTISILVTGGAGFIGSHVYRTLLGPRGVDPVTVTVLIDKLTAYG
TITNLAIEVSDSRFRFRVGDIDAPLVDLDDLLAVHQVVFHFAAEHSHVDRSILGAADFVR
TNVGTOTLLDAALRGQIETFVHIIDTEVIGSIDAGSWPETAAPVSPNSLYSAAKASD
LVALAVHRTGHLDVTRCSNNYGHOPPEKVIPLFVTSLLDGREVLVYGGTVNRDW
LHVDHVRALIEVTRGAGEVYNTGGTELSNKELTQLLLDAGAGHWRVYVYTDK
GHDRIYSDCTKIRBELGYRPAREFGDALAEIVAWYRHRRAWEPFLIRAYCAVAA"
complement(2604..3365)
/gene="mtmrii"
complement(2604..3365)
/gene="mtmrii"
/codon_start=1
/transl_table=11
/product="ketoreductase"
/protein_id="CAA07756.1"
/db_xref="GI:3334821"
/db_xref="GOA:O86480"
/db_xref="SPTREMBL:O86480"
/translation="MDMGLTGRRLVLTGGSSIGIAAVADEGARVALTVRGDEA
ARLELAGAGHAAVARYSIGEPTVESAVATVTERWGGVDVLVANAVWAPRRAP
GTRFEVASGDRNGFLDNLAPTLRIVQLVYAGMEARSGRIVLISSHVLDLCHRGQE
FYGAASLDFARSLAMDVGRGVLVNVVCPGLTITRVLTLGLPDEIRERELGSIPT
GRLSPEDTAGAVFLGSAANANTTGGTLTVSGGR"
complement(3382..4977)
/gene="mtm011"
complement(3382..4977)
/gene="mtm011"
/codon_start=1
/transl_table=11
/product="oxygenase"
/protein_id="CAA07757.1"
/db_xref="GI:3334822"
/db_xref="GOA:O86481"
/db_xref="SPTREMBL:O86481"
/translation="WADPTPLVAGAGVGLTAHRLARGILVRLIDAAPRAATSR
AVATHPRTLETVDQVIEELIENLAQQIRAFNFSRGRRLVRLDADYTIMTRPFFTA
IVQVDTEAVLSAVARUGDVENETRLTGFSDAGVDVTLLEHADGITESTRVVPLV
GCGHSDTTRKLLGPLLIGESSETWMLADAPVDLPRDSIYVHTGQALMVPPFR
GGRWLLDTTAPGTDAAIADIRFTRNSAPASVRQVPTPTSVFTFOQRMVPRMG
EGRVFVAGDAKATFVPSKGRMNTGVQEAENLAWLVAEGHARELLDSYSLERVP
IGERLGSNKKATFLVQKNAPAPIALPVFGLIRNVPPLREKVKORVLGIGSLQLD
YDPSPLTRPAATGAGRPCTRASVSGSDPADPGCRANABELRDTFWLANTVPDVAAG
APAAANWLSVRLIGDSQSPGLADPNGLRLAUGLASGGVLLVDPDGTVAARGALG
GADRLAALAAGLRDEQHVLPHS"
complement(4970..5269)
/gene="mtm011"
complement(4970..5269)
/gene="mtm011"
/codon_start=1
/transl_table=11
/product="oxygenase"
/protein_id="CAA07758.1"
/db_xref="GI:3334823"
/db_xref="SPTREMBL:O86482"
/translation="WYTFVNVKLTVNGDIEEFLTIKESLTERMSAOPGYRSSTTLRQVG
RDPVFTLEAEMDDAASHQAARSEEPQARVKGILGALAVDPGLYEASEDGNRG"
complement(5313..6857)
/gene="mtmL"
complement(5313..6857)
/gene="mtmL"
/codon_start=1
/transl_table=11
/product="acyl CoA ligase"
/protein_id="CAA07759.1"
/db_xref="GI:3334824"
/db_xref="GOA:O86483"
/db_xref="SPTREMBL:O86483"
/translation="MSELPGPAVPLDGLRLLAABERHPQVAVRDERGAVDVAGLDA
ADRFALLRLTGRTPAVGASVLDPVFAAFYRSRGNRVLVNPLVREPVLHV
FTAGIEIALVNSETSPGGBRALPLREVVVDADPGPPPTDLDDLLITDGT
AALPDPAVDLSDVVCVQTSITGTPKGVRLIHRNLVANAQAHAHLGDLGAGVCLN
HLPLVHMLDLSAVYAGATQVLCHDPPDVASAAABAGATHYGLFVLRLARLADR
LASVPGPELGRSGSRLAPAAARLGVPIQGYGLAESLSPSHNDRPERSK
PGSVGPAVPGTECRIVDLETAALDPGRPEVLLRGPQLMAGYLGLPDAPDIAAGWF

```

```

HTDGVGYQDEBGMFLVLDRIKOVFKVDNELVSPSEIEQVLLQDPDVCVADLPDEF
SGAVWAGVVVPGDGPVDLNPVARANALLSDHQRIRRAERLTAVPRSPNGKTERTL
RELRRA"
complement(6850..7797)
/gene="mtmQ"
complement(6850..7797)
/gene="mtmQ"
/codon_start=1
/transl_table=11
/product="aromatase"
/protein_id="CAA07760.1"
/db_xref="GI:3334825"
/db_xref="GOA:Q53669"
/db_xref="SPTREMBL:Q53669"
/translation="MENSQOYSVRDEILVAAPARVRSVLVDVVGWTLHGPAVHAHV
LRDTGDERLIRHWATVRGASRVTRTARHTRSADARSISPAHVMPPEAPFASLEGATFR
ERDCKTRVENRHDFTLSDGASPADLTAIRLREGSRAYLSLTGLHAENHEELRLIIS
FTDSTFAGVDKADYVLEADKWEERIPIHIAKLDLEBPAGVQVDFDMLTSTPDGSSH
RTSSVTVTPEDLIIVKQTELPALLDGHGTGHKFTRTPEGVIAEASHTATVKPSALHI
LGSGTTVQDARNYLRRVLSNSVSNLRKAKAYAEERAGV"
8009..8461
/gene="mtmX"
8009..8461
/gene="mtmX"
/codon_start=1
/transl_table=11
/product="cyclase"
/protein_id="CAA07761.1"
/db_xref="GI:3334826"

```

Query Match 32.7%; Score 163.4; DB 1; Length 18977;  
Best Local Similarity 66.8%; Pred. No. 1.1e-12;  
Matches 233; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY	151	CGGGGGACGTACAGGATCCACTGTCGCCGACGGCGGGAACCTCTCTCTCGCCATG	210
DB	15183	CGCGGGTGGCGAGGATGAACCGTCGCCCTCCGCCAGGAACGGCCCTCTTGGCGAGG	15242
QY	211	ATCTCGTCGGCGTGGTTCAGGGCGAAGCAGCGCGTAGTCCACCGCGTCGGCGGTCAAC	270
DB	15243	ATCTCTCGCGGTAGTTCAGGGCGACGACGAGTAGTAGTGGCTGGTCTTGGCG	15302
QY	271	CGCTCCGGGGTGCACACGGGATGTGCTGCGGGGGTGAGCGGCCCTCTCTTGGCGCGC	330
DB	15303	TACTCGGGGTCGCGACCGGATGTGCTGCGGGCGACGACCGAGCCGCTCTTGAATCG	15362
QY	331	GTCTGTGCGCACACCCAGGAGACGAGTCCGCGACCGATGCCGAGAAAGTTCTGCACCGTG	390
DB	15363	GTGGTGTGCGAGCAGATTCCAGTCTGTCGGGTGAGCCGCGACAGCTTCAGCAGGGTG	15422
QY	391	CGCTCTTTCGCGCTCGCGCCGTACGCCACCAACCGCTTGGCTTGGCTTGGAGCGAGTTC	450
DB	15423	TTCCTTTCGCGCGCGCCCGTACCCGCGATCCGTTTGGCTTGGCTTGGCGCGCAGCGG	15482
QY	451	AGCAGGCGAGCAGTGGTGGGATGCCCTCGACGTCGGCGGCGAACC	499
DB	15483	ACCATGAGGTGAGTTCGGCGGTGATCTCCGCGACGTGCCTGGCGAACC	15531

RESULT 11  
AR088311/c

LOCUS	AR088311	1401 bp	DNA	linear	PAT 07-SEP-2000
DEFINITION	Sequence 1 from patent US 5989869.				
ACCESSION	AR088311				
VERSION	AR088311.1	GI:10015074			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 1401)				
AUTHORS	Filipinski, S., Lomovskaya, N., Fonstein, L., Colombo, A., Luisa. and Hutchinson, C. Richard.				
TITLE	Process for preparing daunorubicin and doxorubicin				



```
JOURNAL Patent: US 5989869-A 1 23-NOV-1999;
FEATURES Location/Qualifiers
source 1..1401
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 30.0%; Score 149.8; DB 6; Length 1401;
Best Local Similarity 65.3%; Pred. No. 1.3e-10;
Matches 220; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 164 GATCCACTGTCGCGACCGCGGAGTCCCTGCTTCCTGCGCATGATCTCGCGGT 223
DB 1390 GGATCCAGCGCCCTCTGCTTCGTGAATCCCGCTCTCGGGCTGAGCTTCGTAGGT 1331
QY 224 GTTTCAGGCGAAGACGAGCGCGTAGTCACCGCGTCGGCGTAAACGCGTCCGGGGTGC 283
DB 1330 GTTTCAGGCGAAGACGAGCGCGTAGTCGGGGTAGGGGGCCCGGAACTCCTCGCGGAGC 1271
QY 284 GCACCGGATGTGCTGCGGGGTGAGCGGGCTCTTGGCGGGGTGCTCGCACA 343
DB 1270 GGATGGGGATGTGGACCCCGGGAGCGACGCGCGATCTTGGCCGGGTGGTGTCTGAGA 1211
QY 344 CCCAGGAGACGAGTCCGCGACCGATGCGCGAGAAAGTTTCGTCACGTCGCGCTCTTCGCCG 403
DB 1210 CGCAGCGCAGACATCGGGGCGGATCCCGCATAGTTGAGCACTGTGCGCTCTTGGCGG 1151
QY 404 TCGGCGCGTACGCCACACCGCTTCGCTCGGCTTGGCGAGTTGAGCGAGTTGAGCAGGCGGACCA 463
DB 1150 TCGGCGCGTACGCCACACCGCTTCGCTCGGCTTGGCGAGTTGAGCGAGTTGAGCAGGCGGACCA 1091
QY 464 GGTGCGTCCGATGCGCTTCGACGCGTTCGCGGCGGAACTT 500
DB 1090 GGTACACAGACCGCGTTGACCTGGGCGCGAACTT 1054

RESULT 12
LOCUS AR144738 1401 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 4 from patent US 6210930.
ACCESSION AR144738
VERSION AR144738.1 GI:15106605
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1401)
AUTHORS Filippini,S., Lomovskaya,N., Fonstein,L., Colombo,A.,Luisa.,
Hutchinson,S.,Richard., Otten,S.L. and Breme,U.
TITLE Process for preparing doxorubicin
JOURNAL Patent: US 6210930-A 4 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1401
/organism="unknown"
/mol_type="unassigned DNA"

Query Match 30.0%; Score 149.8; DB 6; Length 1401;
Best Local Similarity 65.3%; Pred. No. 1.3e-10;
Matches 220; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 164 GATCCACTGTCGCGACCGCGGAGTCCCTGCTTCCTGCGCATGATCTCGCGGT 223
DB 1390 GGATCCAGCGCCCTCTGCTTCGTGAATCCCGCTCTCGGGCTGAGCTTCGTAGGT 1331
QY 224 GTTTCAGGCGAAGACGAGCGCGTAGTCACCGCGTCGGCGTAAACGCGTCCGGGGTGC 283
DB 1330 GTTTCAGGCGAAGACGAGCGCGTAGTCGGGGTAGGGGGCCCGGAACTCCTCGCGGAGC 1271
QY 284 GCACCGGATGTGCTGCGGGGTGAGCGGGCTCTTGGCGGGGTGCTCGCACA 343
DB 1270 GGATGGGGATGTGGACCCCGGGAGCGACGCGCGATCTTGGCCGGGTGGTGTCTGAGA 1211

RESULT 13
LOCUS AF048833/c 2221 bp DNA linear BCT 06-MAY-1998
DEFINITION Streptomyces peucetius polyketide cyclase (dpsI) and daunorubicin/doxorubicin biosynthesis enzyme (dnrX) genes, complete cds.
ACCESSION AF048833
VERSION AF048833.1 GI:2935430
KEYWORDS Streptomyces peucetius
ORGANISM Streptomyces peucetius
REFERENCE 1 (bases 1 to 2221)
AUTHORS Lomovskaya,N., Doi-Katayama,Y., Filippini,S., Nastro,C., Fonstein,L., Gallo,M., Colombo,A.L. and Hutchinson,C.R.
TITLE The streptomyces peucetius dpsI and dnrX genes govern early and late steps of daunorubicin and doxorubicin biosynthesis
JOURNAL J. Bacteriol. 180 (9), 2379-2386 (1998)
MEDLINE 98233733
PUBMED 9573189
REFERENCE 2 (bases 1 to 2221)
AUTHORS Lomovskaya,N., Doi-Katayama,Y., Filippini,S., Nastro,C., Fonstein,L., Gallo,M., Colombo,A.L. and Hutchinson,C.R.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-1998) School of Pharmacy, University of Wisconsin, 425 N. Charter St., Madison, WI 53706, USA
FEATURES Location/Qualifiers
source 1..2221
/organism="Streptomyces peucetius"
/mol_type="genomic DNA"
/strain="ATCC29050"
/db_xref="AFCC:29050"
/db_xref="taxon:1950"
/gene="dpsY"
/gene="dpsY"
/gene="dpsY"
/feature="dpsY; component of daunorubicin/doxorubicin polyketide synthase"
/codon_start=1
/transl_table=11
/product="polyketide cyclase"
/protein_id="AAC38443.1"
/db_xref="GI:2935431"
/translation="MCCGPGAGATWRIIDISSAVDASGWEPEDEVEHVSREGA VMSSEMRHFGVAFDPDELPEGEFLSLDRITLTSHTHIDAPSHYSGRAHYGGRP RNIDELPMFYGGLLDLTCGDGTAGAGLEKELARIGRVPFGTIVLURTGASE RAGTEQYFTDFTGLDPAVNLIDHGVIRVIGTDFSLDAPFGAVRRVRETDGRSLW PAHVTGRHREYQIERLGNLAALPGCDGPFQACFPFKITGGGAGWTRAVAFVDE"
/gene="dnrX"
/gene="dnrX"
/gene="dnrX"
/feature="dnrX"
/codon_start=1
/transl_table=11
/product="daunorubicin/doxorubicin biosynthesis enzyme"
/protein_id="AAC38444.1"
/db_xref="GI:2935432"
```



/translation="MEPNSTCRICGRRVREFFDGRQPLSDYFSPSEELDNEFFRL  
AVGCMVCTMVLLEVRDRMFRYDYPYHSSSRMRHEFAATARRLIGELTGDRP  
FCVEIGSDNVMRLTRVDAGVRHLGVPSGGVADVRAGELQVTRFAFEESTAREIAQ  
EHGPANVIYAANTICHI PYLDSVFRGIDALLAPDGVFVEDPILGDI VEKNTFQIYD  
BHPYLFARSVITTAHQFGEIVLVDVERLPHVHGVEVYTIARAGRQPSRVEGLAER  
SRGLADLTLEKFGQVKVCCDLVARIELRDLGVVGVGTAKATVNLVAGIG  
PLLCCFVYDTTAPKIGRRPLGSHPIRISABEFRAPIFYDIALFWHHLDEVQAREAF  
TKQGRWRWNSG"

## ORIGIN

Query Match 30.0%; Score 149.8; DB 1; Length 2221;  
Best Local Similarity 65.3%; Pred. No. 1.1e-10;  
Matches 220; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 164 GATCCTACCTGTCGCCAGCCGCGCGAATCTCTGCTCCCTCCCATGATCTCGTCGGGT 223  
DB 2210 GGATCAGGGGCCCCCTGCTTGTGAATCCCGCCCTCTCGGCGCTGATCTAGT 2151

QY 224 GGTTCAGGCGAAGACAGCGGTAGTCCACCGCTCGGGCGTGAACGGTCCGGGTGC 283  
DB 2150 GGTTCAGGCGAAGACAGCGCATAGTCGGGTAGGGGCCCCGGAACTCTCTCGGCGAGC 2091

QY 284 GCACCGGATGTCGTCGCGGGGTGAGCCGCGCCCTGCTTGCCCGCGCTGCTGTCGACA 343  
DB 2090 GGATGGGATGTGGACCCGCGGAGCCGACCGCGCATCTTGGCCGCGTGTGCTGATA 2031

QY 344 CCACGAGACCGAGTCCGACCGATGCCCGCAGAGTTCGTCAGTGGCGCTCTCGCG 403  
DB 2030 CGACCGCAGCATCGGGCCCATCCCCCGCATAGTGTAGCATGCTGCGGCTCTTGGCGG 1971

QY 404 TCGCGCGTACGCCACACCCGCTTGCCTTCGCGCTTGAGCGAGTTCAGACGGCGAGCA 463  
DB 1970 TCGCCCGTACCCGACGACGTAGAACCGGAGTTCGCGCAGCTCGCGACGCGGCCACCA 1911

QY 464 GGTCTGTCGGATGCCCTCGACCTCGCGCGGACCT 500  
DB 1910 GGTTCAGCAGACCCGCTTGACCTGGCGCGCAACT 1874

## RESULT 14

BD251846/c  
LOCUS  
DEFINITION  
Gene cluster involved in nogalamin biosynthesis, and its use in production of antibiotics.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BD251846.1 GI:33061616  
JP 2002528068-A/1.  
unidentified  
unclassified.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1. (bases 1 to 16020)  
Ylihanko, K., Torkkeli, S., Palmu, K. and Hakala, J.  
Gene cluster involved in nogalamin biosynthesis, and its use in production of antibiotics  
Patent: JP 2002528068-A 1 03-SEP-2002;  
GALLIAEUS OY

COMMENT  
OS Streptomyces nogalater ATCC 27451

PN JP 2002528068-A/1  
PD 03-SEP-2002  
PF 20-OCT-1999 JP 2000578345  
PR 23-OCT-1998 FI 982295  
PI KRISTINA YLIHONKO, SIRKE TORKKELI, KAISA PALMU, JUHA HAKALA PC  
C12N15/09, C12P7/66, C12P19/56//C07H15/252, (C12P7/66, C12P1.465), PC  
(C12P19/56, C12P1.465), C12N15/00  
CC 'overlapping sequence in the genes snom and snog' CC  
region' 'overlapping sequence in the genes snog and snog' CC 'unknown

FT misc feature 3799..3800  
FT misc feature 6334..6356  
FT misc feature 13201..13300.

## FEATURES

source

Location/Qualifiers  
1..16020

/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

## ORIGIN

Query Match 27.0%; Score 135; DB 6; Length 16020;  
Best Local Similarity 61.5%; Pred. No. 5.4e-09;  
Matches 216; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 150 GCGGGGACGTACAGATCCACTGTCCGCCAGCCCGCGGAACTCTCTGCTTCCTTGCACAT 209  
DB 6323 GCGGGGATGGGCACGATGAACCGCGCGCTGTGTCGCCGAGGCCCTCTCTTGTGCGAG 6264

QY 210 GATCTCTGCGCTGCTTCCAGCGAGACAGCGGTAGTCCACCCGCTCGGGCGGAA 269  
DB 6263 GATCTCTGCGCTGCTTCCAGCGAGACAGGTAGTTCGCGGATGTTCTCTTGGC 6204

QY 270 CGCTCCGGGTGCGCACCGCGGTGTCGCTCCCGGGGTGAGCGCGGCCCTCTGCTTGGCCGG 329  
DB 6203 GTGCTCGGAGCTGACCGGTATGTGGTGGCGGCGAGCACCTGCTGCTTTCAGCTC 6144

QY 330 GTCGTGTGACACACCGAGAGACAGTCCGGACCGATCCGCGAGAGTTCGTACGGT 389  
DB 6143 GGTGCTGCTGCGTATTCAGCTCTTTCAGCGCGAGCCGACACCGGTGAGATGGT 6084

QY 390 GGCGCTCTTCCGCTGCGCGGTACGCCACACCGCTTCCCTCGCGCTTTCAGCGAGTT 449  
DB 6083 GTTGCCCTTGGCGGACACCGTACCGACGATGCGCTTGGCGGACGAGGAGCG 6024

QY 450 CAGCAGGGCAGAGTGGTGGGATGCTCCCTCGACGTCCGGCGCGAACCT 500  
DB 6023 CAGCAGTTCGGCAGTTCGGTGGCGACCCGCTCGATCCGCTGCGCGAACCT 5973

## RESULT 15

AF187532  
LOCUS  
DEFINITION

AF187532 16065 bp DNA linear BCT 07-JUL-2000  
Streptomyces nogalater SnON (snON) gene, partial cds; C-7  
ketoreductase (snof), SnO (snO), SnOL (snOL), putative glycosyl  
transferase (snogE), putative hydroxylase (snomA), putative  
glycosyl transferase (snogD), SnOK (snOK), nogalonic acid methyl  
ester cyclase (snol), putative dTDP-glucose-4-6-dehydratase  
(snogK), putative dTDP-4-dehydroxamose reductase (snogC), SnOG  
(snog), SnOGN (snogN), putative polyketide cyclase (snomA),  
putative amino methylase (snogA), and putative dTDP-glucose  
synthase (snogJ) genes, complete cds; and putative aminotransferase  
(snogI) gene, partial cds.

## ACCESSION

VERSION  
KEYWORDS  
SOURCE

AF187532.1 GI:6018298  
Streptomyces nogalater  
Streptomyces nogalater  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 16065)  
Torkkeli, S., Ylihanko, K., Hakala, J., Skurnik, M. and Mantsala, P.  
Characterization of Streptomyces nogalater genes encoding enzymes  
involved in glycosylation steps in nogalamin biosynthesis  
Mol. Gen. Genet. 256 (2), 203-209 (1997)  
9349712  
PUBMED  
2 (bases 1 to 16065)  
Torkkeli, S., Kunnari, T., Palmu, K., Hakala, J., Mantsala, P. and  
Ylihanko, K.  
Identification of a cyclase gene dictating the C-9 stereochemistry  
of anthracyclines from Streptomyces nogalater  
Antimicrob. Agents Chemother. 44 (2), 396-399 (2000)

## REFERENCE

AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 16065)  
Torkkeli, S., Ylihanko, K., Hakala, J., Skurnik, M. and Mantsala, P.  
Characterization of Streptomyces nogalater genes encoding enzymes  
involved in glycosylation steps in nogalamin biosynthesis  
Mol. Gen. Genet. 256 (2), 203-209 (1997)  
9349712  
PUBMED  
2 (bases 1 to 16065)  
Torkkeli, S., Kunnari, T., Palmu, K., Hakala, J., Mantsala, P. and  
Ylihanko, K.  
Identification of a cyclase gene dictating the C-9 stereochemistry  
of anthracyclines from Streptomyces nogalater  
Antimicrob. Agents Chemother. 44 (2), 396-399 (2000)

## JOURNAL

MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 16065)  
Torkkeli, S., Ylihanko, K., Hakala, J., Skurnik, M. and Mantsala, P.  
Characterization of Streptomyces nogalater genes encoding enzymes  
involved in glycosylation steps in nogalamin biosynthesis  
Mol. Gen. Genet. 256 (2), 203-209 (1997)  
9349712  
PUBMED  
2 (bases 1 to 16065)  
Torkkeli, S., Kunnari, T., Palmu, K., Hakala, J., Mantsala, P. and  
Ylihanko, K.  
Identification of a cyclase gene dictating the C-9 stereochemistry  
of anthracyclines from Streptomyces nogalater  
Antimicrob. Agents Chemother. 44 (2), 396-399 (2000)

## JOURNAL

MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 16065)  
Torkkeli, S., Ylihanko, K., Hakala, J., Skurnik, M. and Mantsala, P.  
Characterization of Streptomyces nogalater genes encoding enzymes  
involved in glycosylation steps in nogalamin biosynthesis  
Mol. Gen. Genet. 256 (2), 203-209 (1997)  
9349712  
PUBMED  
2 (bases 1 to 16065)  
Torkkeli, S., Kunnari, T., Palmu, K., Hakala, J., Mantsala, P. and  
Ylihanko, K.  
Identification of a cyclase gene dictating the C-9 stereochemistry  
of anthracyclines from Streptomyces nogalater  
Antimicrob. Agents Chemother. 44 (2), 396-399 (2000)

## JOURNAL

MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 16065)  
Torkkeli, S., Ylihanko, K., Hakala, J., Skurnik, M. and Mantsala, P.  
Characterization of Streptomyces nogalater genes encoding enzymes  
involved in glycosylation steps in nogalamin biosynthesis  
Mol. Gen. Genet. 256 (2), 203-209 (1997)  
9349712  
PUBMED  
2 (bases 1 to 16065)  
Torkkeli, S., Kunnari, T., Palmu, K., Hakala, J., Mantsala, P. and  
Ylihanko, K.  
Identification of a cyclase gene dictating the C-9 stereochemistry  
of anthracyclines from Streptomyces nogalater  
Antimicrob. Agents Chemother. 44 (2), 396-399 (2000)

## JOURNAL

MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 16065)  
Torkkeli, S., Ylihanko, K., Hakala, J., Skurnik, M. and Mantsala, P.  
Characterization of Streptomyces nogalater genes encoding enzymes  
involved in glycosylation steps in nogalamin biosynthesis  
Mol. Gen. Genet. 256 (2), 203-209 (1997)  
9349712  
PUBMED  
2 (bases 1 to 16065)  
Torkkeli, S., Kunnari, T., Palmu, K., Hakala, J., Mantsala, P. and  
Ylihanko, K.  
Identification of a cyclase gene dictating the C-9 stereochemistry  
of anthracyclines from Streptomyces nogalater  
Antimicrob. Agents Chemother. 44 (2), 396-399 (2000)

## JOURNAL

MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 16065)  
Torkkeli, S., Ylihanko, K., Hakala, J., Skurnik, M. and Mantsala, P.  
Characterization of Streptomyces nogalater genes encoding enzymes  
involved in glycosylation steps in nogalamin biosynthesis  
Mol. Gen. Genet. 256 (2), 203-209 (1997)  
9349712  
PUBMED  
2 (bases 1 to 16065)  
Torkkeli, S., Kunnari, T., Palmu, K., Hakala, J., Mantsala, P. and  
Ylihanko, K.  
Identification of a cyclase gene dictating the C-9 stereochemistry  
of anthracyclines from Streptomyces nogalater  
Antimicrob. Agents Chemother. 44 (2), 396-399 (2000)

Vatselankatu 2, Turku FIN-20014, Finland  
 Location/Qualifiers  
 1. 16065  
 /organism="Streptomyces nogalater"  
 /mol\_type="genomic DNA"  
 /strain="ATCC27451"  
 /db\_xref="ATCC:27451"  
 /note="related downstream sequence in EMBL accession number AJ224512"  
 complement (<1..571)  
 /gene="snoN"  
 complement (<1..571)  
 /gene="snoN"  
 /codon\_start=1  
 /transl\_table=11  
 /product="SnoN"  
 /protein\_id="AA01805.1"  
 /db\_xref="GI:6018299"  
 /translation="MQTEPGVADLPAESDPALRELAARVERDGVVPGVLDAE  
 VAEYLAARLLAHEESVRNMGSGAGTVMYDAQAQSGDMRELAHPRIAAAEVLA  
 GSPRLFKLEVLKENKQASVPTAPHDFAFPFFSTAGTALTAWALVDVPEVREGC  
 MTFVPGSHLLPDDTGDPEWAGAFTRPGEI"  
 complement (644..1489)  
 /gene="snoaf"  
 complement (644..1489)  
 /gene="snoaf"  
 /codon\_start=1  
 /transl\_table=11  
 /product="C-7 ketoreductase"  
 /protein\_id="AA01806.1"  
 /db\_xref="GI:6018300"  
 /translation="PMRATDSTGRPVAMPSPAPSPPTSPAPGSPAPLAVITGG  
 CSGGRATAPAAQAKVAVVGTEDALQAQAGCADMLVADVASPDGPAQVNA  
 ALRFGTRIDVUNNAVAGMETLOTYDRDAVARQFQNLAPLPLVQSALGALKESG  
 IVNVGTAAATGLRAAPATGALYGASKVLDLITINAVELAPGRIRVGVVAGVIDTG  
 IGVKGMTPEGRBFLTMGGRVPGVRGVEDVAMWIVQLARFEAGTATGMVVFVVG  
 GLSLV"  
 complement (1580..2047)  
 /gene="snoO"  
 complement (1580..2047)  
 /gene="snoO"  
 /codon\_start=1  
 /transl\_table=11  
 /product="SnoO"  
 /protein\_id="AA01807.1"  
 /db\_xref="GI:6018301"  
 /translation="MSVTDQTAAPEDRAAATDPGFHLYAQVOQFYAROMQLDSCG  
 AEWAAFTEDGTAPRSPSPAPRCHAEALAGARAAERLAEGLSHEHVIGMTAVRR  
 EPDGSVFRSYAQVATRRGGEAPRLHLICVCEDVLVREGPGIKVRVRVTHDA"  
 2267..2686  
 /gene="snoL"  
 2267..2686  
 /gene="snoL"  
 /codon\_start=1  
 /transl\_table=11  
 /product="SnoL"  
 /protein\_id="AA01808.1"  
 /db\_xref="GI:6018302"  
 /translation="MSTANKERCLEWMAANRWDSVGVAAHAPDVVHYDDEKPV  
 AEEVVRMSAVEAPDLRLDVSIVGEGDVRMLRITCSATHQGVFMGIAFTGRKVR  
 TVLEELRFEAGVVEHVDVFNFSPLFRDLGVVPDGL"  
 complement (2744..4042)  
 /gene="snoG"  
 complement (2744..4042)  
 /gene="snoG"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative glycosyl transferase"  
 /protein\_id="AA01809.1"  
 /db\_xref="GI:6018303"  
 /translation="MRVLLTSFAMDAHCTAVPLAWLASAGHEVVRVACQALPTTIT  
 GAGLTAVPVGRDTHGSLLRGDLALHDEADYLEARHDALEFELKGHNTVMSAL

FYSIINDSMVDLIDFARHWRPDLVWVEPTEFAGAVARASGAHAARLLSFDPDLFUS  
 TRRLFLERMARQBEHDDTLAWLDMTLGRGHGSHDEEIVTGOWSIDOTPAVPLDA  
 GGTPVMRYVPSGLVFTVVDVWMLRPERPRVLTGITSRSVKSLAVSVDLFEA  
 VAGLVQVATLADQRELGRVDPDHRIVEHVPLDVAFTCSAIVHGGAGTWTSTAA  
 VYGVQVSLGSMMDHFTYARLEBELGAGLRPLSGELTAGRLTRELRLVIGEPSPGTAA  
 QALSDTIAAPSPSEVVVLEELTGRHPCGTREPRALRA"  
 complement (4086..4913)  
 /gene="snoaw"  
 complement (4086..4913)  
 /gene="snoaw"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative hydroxylase"  
 /protein\_id="AA01810.1"  
 /db\_xref="GI:6018304"  
 /translation="MTVLVTGATGNVGRHVHTVGLLAARRVRALTRTPDRSGLPCEGAE  
 ITGDLTRPEYERMLDQVEAVYLPVPEATAAAGARAGVRIIVLSSDSVTDTGT  
 ITGGHRRVELAVEDTGLEHWVRFEGFALNKVITWASIRAEGVRSAYPARVAPVH  
 EADVAVAALLKKEGGRAYSVTGQALTRQEQVRAVGEGLSRLAFVETVQGAR  
 ADLTAQGLPAPADYVLAFOAGWTERPAPAPTVREVTGRTPARTLAQMAADHRADRFR"  
 5053..6225  
 /gene="snoGD"  
 5053..6225  
 /gene="snoGD"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative glycosyl transferase"  
 /protein\_id="AA01811.1"  
 /db\_xref="GI:6018305"  
 /translation="MRVPSGSCRTGIMRALPITPSGLSHILPTVLAQALRALGHEVR  
 YATGDIRAVABAGLCAYDVSGVNYAKLFVDDTDVTDPMHSEGLSGGFFFAEMFARV  
 SAVADGALRTARSRRPDLVWVTPQAGAPLTAALQIPCELPGLPADSERGICALI  
 REAMVGGRTVGRHVGTGTPSVRLTTPPSVALLPELDRSGAMWPMRYVYNGGAVLP  
 DWLPAAQRRIAVTIGSIDALSGGIKAPLAFSEVADVAFVLTGGLDGLALGEL  
 PANVRVWVPIGLALLETCDALIHGGSGTLTALAAGVQACVIPHGSYQDTNRDLT  
 GLGIGFDABAGSLGAEQCRRLDDAGLREALLRVRQEMSEMPPEAETAAKLVALAG"  
 complement (6271..7074)  
 /gene="snoK"  
 complement (6271..7074)  
 /gene="snoK"  
 /codon\_start=1  
 /transl\_table=11  
 /product="SnoK"  
 /protein\_id="AA01812.1"  
 /db\_xref="GI:6018306"  
 /translation="MPDFGPTAENLSKEAVRFVREQYVHIPLVLSSTEVTAFAAA  
 CBEVKEGREIWAGEDVQVHYVAQWQKPELSLVHPEISGLARLAGAPLRY  
 YSDILVKEPKRTPLTHDDDETGLPLNELSATLTAWIALTDVPIVERGCMQSVPGSHL  
 RAREDRQEMTSFAEFDLADVMPDPYQWQVAVPVAGDVVFFHCHRTVHMAENISD  
 SVMAHGVMYMDADATYRPGVDQGHLSRLSPGDPLEGELPLVTAGTQ"  
 complement (7125..7529)  
 /gene="snoal"  
 complement (7125..7529)  
 /gene="snoal"  
 /codon\_start=1  
 /transl\_table=11  
 /product="nogalonic acid methyl ester cyclase"  
 /protein\_id="AA01813.1"  
 /db\_xref="GI:6018307"  
 /translation="MVSANTGTDDVDVHDPYLNLPATLEHIGITGPKAFQAQLVGM  
 VRAFSEAKLEVRLEERGPPVAKVLYGHEVGLVGMPTDFRFSQEQVHLMRIV  
 DGKIRDRDMPDQGLRLRQLGDPWDEGRWP"  
 7769..8821  
 /gene="snoGK"  
 7769..8821  
 /gene="snoGK"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative dTDP-glucose-4,6-dehydratase"  
 /protein\_id="AA01814.1"  
 /db\_xref="GI:6018308"  
 /translation="MASHTSATDDVILVTGAVGFIGSAYVRMLLENRAPGAGAPVR  
 VTVDLKLTYAGNLNLDAVRGDRLFRVRGIDLAELVDELMASHSDQVHFAESHVDR

STRADADFLVTVVQTRLLDAAALRHGVEPFLVSTDEVYSGSIAGSWPEHPHPLSPNS  
PYAASKAGADLMAFACHRTGLDVRVTRCPNHYGPHPEKLIIRFVFNLLDGLFVPL  
YGDGNVREWLHVEDHCRGVDLVRTAGPFGVYHIGGRELNSRELVMLELGCADW  
SSVRVPDRKQHDRLAYSLDWGRARELGYRPAEPSSGLRSTVQWYRENRSWWEPFLK  
GVTAPGTSVTPGVGR"  
8857..9732  
/gene="snogC"  
8857..9732  
/gene="snogC"  
/codon\_start=1  
/transl\_table=11  
/product="putative dTDP-4-dehydrorhamnose reductase"  
/db\_xref="GI:6018309"  
/translation="MLASHLTAAALFTGSRPEAEAVLGRALDITDGRAVDAAFA  
HPRVNVNCAFTVDGAESRWAEAVRVNGSPRLARCAHGWRLVLLHVSITDVFPG  
DTRSPYGSDAPPTVYGRSKLAGERAVLSLLPDTGTVTVRTHVYGGQSRFVKTML

Query Match 27.0%; Score 135; DB 1; Length 16065;  
Best Local Similarity 61.5%; Pred. No. 5.4e-09;  
Matches 216; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
QY 150 GCGGGGAGTACAGATCACTGTCGGCAGCCGCGGGAACCTGCTCTTCGCCAT 209  
DB 9743 GCGGGGAGTACAGATCACTGTCGGCAGCCGCGGGAACCTGCTCTTCGCCAT 9802  
QY 210 GATCTCGTTCGGGTTCAGCGGAGAGCAGCGGTAGTCCACCGCTCGGGCTGAA 269  
DB 9803 GATCTCGTTCGGGTTCAGCGGAGAGCAGCGGTAGTTCGGGATGTCTTCGGC 9862  
QY 270 CGCGTTCGGGTTCAGCGGAGAGCAGCGGTAGTTCGGGATGTCTTCGGC 329  
DB 9863 GTCTCGGAGCGTTCAGCGGAGAGCAGCGGTAGTTCGGGATGTCTTCGGC 9922  
QY 330 CGTCGTTCGACACCGGAGAGCAGCGGTAGTTCGGGATGTCTTCGGC 389  
DB 9923 GTCTCGGAGCGTTCAGCGGAGAGCAGCGGTAGTTCGGGATGTCTTCGGC 9982  
QY 390 GCGCGCTTCGCGGTTCAGCGGAGAGCAGCGGTAGTTCGGGATGTCTTCGGC 449  
DB 9983 GTCTCGGAGCGTTCAGCGGAGAGCAGCGGTAGTTCGGGATGTCTTCGGC 10042  
QY 450 CAGCAGGCGGAGCGTTCAGCGGAGAGCAGCGGTAGTTCGGGATGTCTTCGGC 500  
DB 10043 CAGCAGTTCGCGGTTCAGCGGAGAGCAGCGGTAGTTCGGGATGTCTTCGGC 10093

RESULT 16  
AX196110  
LOCUS AX196110 12152 bp DNA linear PAT 28-AUG-2001  
DEFINITION Sequence 182 from Patent WO0151639.  
ACCESSION AX196110  
VERSION AX196110.1 GI:15386342  
KEYWORDS  
SOURCE Micromonospora carbonacea  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteridae; Actinomycetales;  
REFERENCE 1  
Hosted, T.J., Horan, A.C. and Wang, T.X.  
Everninomicin biosynthetic genes  
Patent: WO 0151639-A 182 19-JUL-2001;  
Schering Corporation (US)  
FEATURES  
source  
1..12152  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

ORIGIN  
Query Match 26.2%; Score 131; DB 6; Length 12152;  
Best Local Similarity 100.0%; Pred. No. 1.9e-08;  
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTACCCGACCGGTGTCCCGGAAACAAGTTCGAGATACGGCGAGAGAACACCCCGGT 60  
DB 12022 GGTACCCGACCGGTGTCCCGGAAACAAGTTCGAGATACGGCGAGAGAACACCCCGGT 12081  
QY 61 AGTCCGGGTAGACGGTGGGCGGAAAGCGGTACCGGCTTCGACGGTCAGCGGCGGCGG 120  
DB 12082 AGTCCGGGTAGACGGTGGGCGGAAAGCGGTACCGGCTTCGACGGTCAGCGGCGGCGG 12141  
QY 121 ACACCGGCGG 131  
DB 12142 ACACCGGCGG 12152

RESULT 17  
AX574140/c  
LOCUS AX574140 603 bp DNA linear PAT 07-JAN-2003  
DEFINITION Sequence 220 from Patent WO02079505.  
ACCESSION AX574140  
VERSION AX574140.1 GI:27551677  
KEYWORDS  
SOURCE Micromonospora carbonacea  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteridae; Actinomycetales;  
REFERENCE 1  
Farnet, C.M., Zazopoulos, E. and Staffa, A.  
Compositions and methods for identifying and distinguishing  
orthosomycin biosynthetic loci  
Patent: WO 02079505-A 220 10-OCT-2002;  
Ecopia Biosciences Inc. (CA)  
FEATURES  
source  
1..603  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"  
/note="var. africana"

## ORIGIN

Query Match 24.8%; Score 124; DB 6; Length 603;  
Best Local Similarity 100.0%; Pred. No. 3.6e-07;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGTACCCGACCGGTGTCCCGGAAACAAGTTCGAGATACGGCGAGAGAACACCCCGGT 60  
DB 124 GGTACCCGACCGGTGTCCCGGAAACAAGTTCGAGATACGGCGAGAGAACACCCCGGT 65  
QY 61 AGTCCGGGTAGACGGTGGGCGGAAAGCGGTACCGGCTTCGACGGTCAGCGGCGGCGG 120  
DB 64 AGTCCGGGTAGACGGTGGGCGGAAAGCGGTACCGGCTTCGACGGTCAGCGGCGGCGG 5  
QY 121 ACAC 124  
DB 4 ACAC 1

## RESULT 18

AX574198/c  
LOCUS AX574198 15240 bp DNA linear PAT 07-JAN-2003  
DEFINITION Sequence 278 from Patent WO02079505.  
ACCESSION AX574198  
VERSION AX574198.1 GI:27551677  
KEYWORDS  
SOURCE Micromonospora carbonacea  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteridae; Actinomycetales;  
REFERENCE 1  
Farnet, C.M., Zazopoulos, E. and Staffa, A.  
Compositions and methods for identifying and distinguishing  
orthosomycin biosynthetic loci  
Patent: WO 02079505-A 278 10-OCT-2002;  
Ecopia Biosciences Inc. (CA)

## ORIGIN

Query Match 26.2%; Score 131; DB 6; Length 12152;  
Best Local Similarity 100.0%; Pred. No. 1.9e-08;  
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEATURES Location/Qualifiers  
source 1..15240  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"  
/note="var. africana"

ORIGIN  
Query Match 23.6%; Score 118.2; DB 6; Length 15240;  
Best Local Similarity 53.3%; Pred. No. 8.3e-07;  
Matches 249; Conservative 0; Mismatches 218; Indels 0; Gaps 0;  
QY 33 AGATACGGCGAGAGAACCCCGGTACTCGGGTAGAGTGGCGGCGAAGGCGTAC 92  
Db 12684 AGAGATCGATCCAGCGCAGCAGCACATGTCGCGGACCGGGCTCCGGGTACATGA 12625  
QY 93 GCGCTTCGACGGTTCAGCGCGCGCGGACACCGCGGGTCACTGTCACGTGACGG 152  
Db 12624 AAGTTGATACCTCGGTGGGGCGGTGTATCCCGTCCGCGACCGTCTCGACCG 12565  
QY 153 GGGACGCTACAGGATCCACTGTCGCGCAGCCCGCGGAATCTCTGCTTCCTTCGCCATGAT 212  
Db 12564 GGGCAGCGGAAGATCAACCGGGTGGCGTGGCCAGCATCTCCGCTCCCGGCGCAGAT 12505  
QY 213 CTCGCGCGGTGTTCCAGCGAGACGCGGTAGTCCACCGGTCGCGGCGTGAACGC 272  
Db 12504 CTCGTCGCGGAGTCCAGCGCAGCAGGATAGTATGTCGCGGCGCGCCCGGACTC 12445  
QY 273 GTCGCGGGTGCACCGCGGATGTCGTCGCGGGGTGAGCGCGCCCTGCTGTCGCCCGGT 332  
Db 12444 CTGCTCGTATGATTTTCGATGTCGTCGCGAGGTGCGCGGCCACCTTGTCCGGATT 12385  
QY 333 CGTGTGCGACACCCAGAGACGAGTCCGCGACCGATCCCGAGAGTTCCTCAGGTGC 392  
Db 12384 CGCTCCGCGCGTACCGGATGAGCTCGCGTCTATCCGCGAAGTGCAGAGGGTGT 12325  
QY 393 GCTCTTCGCGTCCGCGCGTACCGACACCGCGCTTGCCTTCGCGCTTGCAGCGAGTTCAG 452  
Db 12324 GCCCTTCGTCAGCGCGGTAGAGTGCACGCTCGCGGCGCTTGCCTCGCGAGTTCGCGCAG 12265  
QY 453 CAGGCGAGCGAGTTCGTCGCGATGCGCTCGACGTCGCGCGCGGAAC 499  
Db 12264 CAGGCGGTACCTGTCACGCTGTCGCGACCTGTCGCGGAGC 12218

RESULT 19  
AY131229/c  
LOCUS  
DEFINITION Streptomyces sp. 139 exopolysaccharide gene cluster, complete sequence.  
ACCESSION AY131229 AF538923  
VERSION AY131229.2 GI:22657428  
KEYWORDS  
SOURCE Streptomyces sp. 139  
ORGANISM Streptomyces sp. 139  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 31310)  
AUTHORS Wang, L., Li, S. and Li, Y.  
TITLE Identification and characterization of a new exopolysaccharide biosynthesis gene cluster from Streptomyces  
JOURNAL FEMS Microbiol. Lett. 220 (1), 21-27 (2003)  
MEDLINE 22531099  
PUBMED 12644223  
REFERENCE 2 (bases 1 to 31310)  
AUTHORS Wang, L., Li, S. and Li, Y.  
TITLE Isolation and sequencing of glycosyltransferase gene and UDP-glucose dehydrogenase gene that are located on a gene cluster involved in a new exopolysaccharide biosynthesis in Streptomyces  
JOURNAL DNA Seq. 14 (2), 141-145 (2003)  
REFERENCE 3 (bases 1 to 31310)  
AUTHORS Wang, L., Li, S. and Li, Y.  
TITLE Direct Submission

JOURNAL Submitted (12-JUL-2002) Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences, Tiantan Xili 1, Beijing, Beijing 100050, People's Republic of China  
REFERENCE 4 (bases 1 to 31310)  
AUTHORS Wang, L., Li, S. and Li, Y.  
JOURNAL Direct Submission  
SUBMITTED (03-SEP-2002) Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences, Tiantan Xili 1, Beijing, Beijing 100050, People's Republic of China  
REMARK Sequence update by submitter  
COMMENT On Sep 3, 2002 this sequence version replaced gi:2252974.  
FEATURES Location/Qualifiers  
source 1..31310  
/organism="Streptomyces sp. 139"  
/mol\_type="genomic DNA"  
/strain="139"  
/db\_xref="taxon:203783"  
/clone="cosmid pLV501"  
52..663  
/note="orf1; similar to Streptomyces coelicolor A3 (2) conserved hypothetical protein encoded by GenBank Accession Number AL512944 and to Streptomyces nodosus AmphoRF1 encoded by GenBank Accession Number AF357202"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AA04226.1"  
/db\_xref="GI:22657428"  
/translation="MAGSRVIDYAKDHLHGLRRDRALLKIDGLSEYDARRPLT ATGNLGLVATVATVEARYFEGFRSPQPLRPWDAGSDLWAGGTRDGVGFG YLRWHSDDATDALDTPGRVPMWPEHADTNLFAVMHVLGETVHTGHADILRE GAGTGLRPSHERPIDDEARAAHRAKLERAGAAESPLRSR"  
660..1403  
/note="orf2; similar to Streptomyces coelicolor A3 (2) conserved hypothetical protein encoded by GenBank Accession Number AL58061 and to Bacillus halodurans BH0697 unknown conserved protein encoded by GenBank Accession Number AP001509; pfam03575; peptidase family S51."  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="unknown"  
/protein\_id="AA04227.1"  
/db\_xref="GI:22657430"  
/translation="MTSREPTIVATGGHRTVRVNFDAVHVAVDLSGVGHRPR VNYGTATGDAEHFTARMTEAARAGTADLTPLHLFPMNVDEATVLDQVWVMGG SVANLLAVRVHGLDGMRAWQAGVLSGVSAGSLCWFEAGTDFQPELRPLTNAL GLPYGNGVHDTIDGRRPLRHLVADGTLPTAHTCTDDGVLVGRGTVEAIVEIPG KGAYLVRRGEGDAVEERIEPRMLIVFST"  
780..1173  
/note="similar to pfam03575 and peptidase family S51"  
/complement(1428..2192)  
/gene="stel"  
/complement(1428..2192)  
/gene="stel"  
/note="similar to gntR-family transcriptional regulators, including Streptomyces griseus Daser encoded by GenBank Accession Number AB061860 and Streptomyces coelicolor SC7E4.28 encoded by GenBank Accession Number AL359214; pfam00392"  
/codon\_start=1  
/transl\_table=11  
/product="putative gntR-family transcriptional regulator"  
/protein\_id="AA04228.1"  
/db\_xref="GI:22657431"  
/translation="MSTDVSSAGADTRAPRRDRVPKYLIKORLQMTDERAPGSPM PERRLLAVFQTSRTTVRKALLELSEGLDRIOGKGTFRVAPKVTIQLTSTEDM RAQGLSPASQVLDIGVPADALAELEVEDRVLRIERLAGEPMIAEATHLSA RRPFGRLNRTYLSITVLAQVGVRADEATETSPATPREAGLLGTDLGLPMLL LSRHSDENGAPVFWVRVRSRYKFAIRLRPRG"  
complement(1944..2114)  
misc\_feature  
gene  
CDS

```

/ gene="ste1"
/ note="Region: pfam00392 and smart00345; similar to gntR
family and helix-turn-helix gluconate operon
transcriptional repressor"
/ gene="ste2"
/ gene="ste2"
/ gene="ste2"
/ note="similar to putative sugar transporter sugar-binding
proteins, including Streptomyces coelicolor SC7E4.29
encoded by GenBank Accession Number AL359214 and
Streptomyces griseus Dasa encoded by GenBank Accession
Number AB061860; pfam01547; PS00013; contains possible
N-terminal region signal peptide sequence"
/ codon_start=1
/ translatable=1
/ product="putative sugar transporter sugar-binding
protein"
/ protein_id="AA04229.1"
/ db_xref="GI:22657432"
/ translation="MCAQGIPEEGWVKEKLAALGVAGMLPSVAACGGDDGGKKA
GUDGYAGQILVWMDGSSPDQKDLTAFEFKTKYAKVKEVQQRNGIQKLTALS
ENPFDVIGNTQIPAYAKTGLADLADLKEIGWESLNESSVFEKGQYAAPWY
FANRVLYNKKWADAGLKDTPKTRDEFYALKTIKAKTDAEPIYLPFGQWVHFGVLY
VGGSLVKGDDKYVNLDDPKVAKAETKFKQALSKAPKDEATPQQGEVFAKQ
NVGAFICMGWEAGIALEANPAIEKEIGYFTIPGATADKEPVFLGSLNLAAGSQKQ
ELAKFLKALSQYEGQALKNGVLPNKESLQGNLKNAPAEAPAAVAGTTPLI
PEWAAVENDPNPVKTYLTAVLNGKSPAAAKVQEGFENKRLAQQQ"
/ misc_feature
2560..2595
/ gene="ste2"
/ note="Region: PS00013; similar to prokaryotic membrane
lipoprotein lipid attachment site; positioned
appropriately within protein"
/ misc_feature
2592..3762
/ gene="ste2"
/ note="Region: pfam01547; similar to bacterial
extracellular solute-binding protein"
/ gene="ste3"
3932..4918
/ gene="ste3"
3932..4918
/ gene="ste3"
/ note="similar to putative ABC transporter integral
membrane proteins, including Streptomyces coelicolor
SC7E4.30 encoded by GenBank Accession Number AL359214 and
Streptomyces griseus Dasa encoded by GenBank Accession
Number AB061860; pfam00528"
/ codon_start=1
/ translatable=1
/ product="putative sugar transporter integral membrane
protein"
/ protein_id="AA04230.1"
/ db_xref="GI:22657433"
/ translation="MTVQTERPPSGADVRSKGGTAPGGPPPPAASRAGALAPYLILL
PACATVLLGWPLLDLLSLFQNMALQIQRVTEWNGIDNYKEVLTGEDFVRVTLR
SILFTAVNALVLLGLALVGLLRKMRMTLMIGLVWAMPVVAATTVYQWLFA
QRGVNVLDDLGWMSMADYSWTSSCMSTFFVVTLLIVMWSVPFVILNLYAATTIP
GELYEAALDAGAWNSFTVTLPELRPELYANTELLEIWIWKAFOVTEGGGPDRL
LTSLPVPYAEIGVQHGQMGAAIVLILMLGLTAYLIRLVKQEDAL"
/ misc_feature
4577..4774
/ gene="ste3"
/ note="Region: pfam00528; similar to
binding-protein-dependent transport systems inner membrane
component"
/ gene="ste4"
4915..5748
/ gene="ste4"
4915..5748
/ gene="ste4"
/ note="similar to putative ABC transporter integral
membrane proteins, including Streptomyces coelicolor
SC7E4.31 encoded by GenBank Accession Number AL359214 and
Streptomyces griseus Dasa encoded by GenBank Accession
Number AB061860; pfam00528; PS00013; contains possible
hydrophobic membrane-spanning domains"

```

```

/ codon_start=1
/ translatable=1
/ product="putative sugar transporter integral membrane
protein"
/ protein_id="AA04231.1"
/ db_xref="GI:22657434"
/ translation="MKESLGRVMPNATALVVFVACLPVYVMPATALKPTGDIASD
PVWPTVTEHEFATAVDADHFYTVYRNSLTCTCLAVLFSLGIALAGSPALARMFKG
RRTGIVGFLMAQAPMEVWVIALYIMIVRDSMLNSLVPLTAFTMMMLPPTLILRGP
VAAVPKLESAMVDDCTRGQAFRRVILPLLAGLMTSMFGFITANNEPFLVLVLSK
ESGQAQTLPLWLSSFTQFGDWDGATMAASSLFAVPTLILFVLQRRSAVSGLTAGAVK
G"
/ misc_feature
4948..4983
/ gene="ste4"
/ note="Region: PS00013; similar to prokaryotic membrane
lipoprotein lipid attachment site"
/ gene="ste4"
5320..5616
/ note="Region: pfam00528; similar to
binding-protein-dependent transport systems inner membrane
component"
/ gene="ste5"
6674..8152
/ note="ste5"
6674..8152
/ gene="ste5"
/ note="ste5; possible glycosyl transferase; similar to
glycosyl transferases, including Streptomyces coelicolor"

```

```

Query Match 23.2%; Score 116.2; DB 1; Length 31310;
Best Local Similarity 58.6%; Pred. No. 1.2e-06; Gaps 0;
Matches 202; Conservative 0; Mismatches 143; Indels 0;
QY 154 GGGAGCTACAGGATCCACTGTCCGCCAGCCGCGGGAACCTCTCTCTCTCCCATGATC 213
DB 25124 GGGATGGGAAGACAGCCGCCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25065
QY 214 TCGTCGGCGTGTTCAGGCGAAGAGCAGCCGCTAGTCCACCGCTCGGGCGTGAACGCG 273
DB 25064 TCGGCCCGCAGGTTCCAGGGCAGCACAGCAGTAGTCCGCGCGGCGGCGGCGG 25005
QY 274 TCGGGGTGCGCACCGGATGTGCGTCCGCGGGGTGAGCCGCGCTCTCTCTCTCTCTCT 333
DB 25004 TCGGGCGCAGATCGGATCGGGTGTCCCGGGGTGTACCGCGCTCTCTCTCTCTCT 24945
QY 334 GTGTGCGCACCCAGGAGACAGGTCGGAACCGATGCGCGCAGAGTTCGTCACCGTGGCG 393
DB 24944 CGGTCCACCGTGTACGGCAGCAGGTTCGGCGCGCGCAGCGCAGTGTGTGAGCAGG 24885
QY 394 CTCCTTCGCGGTGCGCGCGTACGCCACACCGCTTGCCTTCGCGCTTGAGCAGTTCAGC 453
DB 24884 CCTTTCGCGGGGCGCGCTACCGACGACCGCTTCTCGCTCTCTCTCTCTCTCTCT 24825
QY 454 AGGCGCAGCAGTGGTTCGGATGCCCTTCGACGTCCGCGCGGGAAC 498
DB 24824 AACCCAGCAGTCCCGCGCACCTTGGCCACCGCGCGGCGGAGAAC 24780

```

## RESULT 20

AX195978/c

LOCUS

Sequence 50 from Patent WO0151639.

DEFINITION

AX195978

VERSION

AX195978.1 GI:15386210

KEYWORDS

SOURCE

ORGANISM

Micromonospora carbonacea

Micromonospora carbonacea

Bacteria; Actinobacteria;

Micromonosporaceae; Micromonospora.

REFERENCE

1

AUTHORS

Hosted, T.J., Horan, A.C. and Wang, T.X.

Eveningomycin biosynthetic genes

Patent: WO 0151639-A 50 19-JUL-2001;

```

FEATURES             Schering Corporation (US)
source               Location/Qualifiers
1..1236
/organism="Micromonospora carbonacea"
/mol_type="unassigned DNA"
/db_xref="taxon:47853"
1..1236
CDS
/notes="unnamed protein product; evrM"
/codon_start=1
/transl_table=11
/protein_id="CAC60015.1"
/db_xref="GI:15386211"
/db_xref="REFSEQ:NM_001001511"
/translation="MVERITTCVCGGTLRTVLDGAQYLGKSVKGVSDPRVRPPL
ELRRCARDGDPGLVQLRHTLPGLLVDTYWRINDTWRHLADYVDAEVLG
ERPKVLIDICGNQDILGNTDAQRWIDPSSAADADPNLIRDFPSPVLPHAG
SFLVTSIMAFYVEDVEFARQVEALLAPGVVYVAVLPALEIAGVDSICHELH
SVYSLATLSILGALGLEIYVRSINGVGGSIICVTVRAGEVGRADGSKALADREL
ALGLDGPAPQRFARQVQRHDEVSALLRELRGQETVHVYGSTKNTLLQFCGIDR
ELRIYAERNPDKVGAATGLTDGIIISQESRAARPDYLVLPWFRDRIIVAREBML
ASGRLTFLPLPRVETVGRPG"

ORIGIN
Query Match      22.6%; Score 112.8; DB 6; Length 1236;
Best Local Similarity 56.5%; Pred. No. 8.4e-06;
Matches 210; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 128 CGGTCAGCTCGTCACGTGTACGCGGGGAGCTACAGATCCACTGTCCGCGACGCCCGC 187
DB 1276 CGGTCGCGCCACCGTCTCGACCCGCGGACGCGGAGATCAACCGGTCGCGTGGCCA 1217
QY 188 GGAATCTCTGCTCTCCATGATCTCGTCGGCGTGTTCAGCGGAGAGCAGCGGT 247
DB 1216 GCATCTCCGCTCCCGGCGAGATCTCGTCCCGGAGTGCAGGCGACGAGGTAGT 1157
QY 248 AGTCACACCGCTCGGCGGTGAACCGCTCCGCGGTGGCGACCGGATGTGCTGCGGGGG 307
DB 1156 AGTCGCGGCGGCGCGCGGCTCTCTCTCGTATGATTTGATGCTGCTGCGAGGG 1097
QY 308 TGAGCGCGCCCTGCTGTGCGCGGTGCTGTCGACACCCAGGAGACAGGTCCGACCGA 367
DB 1096 TSCGCGCGCCACCTGTGTCGCGATTCGCGTCCGCGGCTACCGGATGAGCTCGCGGTCTA 1037
QY 368 TCCCGCAGAGTTCGTACGCTGGCGCTCTTCGCGTCCGCGTCCGCGACCCACCGCT 427
DB 1036 TCCCGCAGAGTTCGAGGAGGTGTTGCCCTTCGTGACGCGCGGTAGACGTGCGCGTGC 977
QY 428 TGCCCTCGGCTTGAGCGAGTTTCAGCGGCGAGCAGGTGCGGTGCGGATGCCCTGACGT 487
DB 976 GGCCTGCGCGCGCGAGCTCGCGACGAGGCGCTCACCTCGTACGCTGCTGCGCGACCT 917
QY 488 CGGCGCGGAACC 499
DB 916 GCTCGCGGAAGC 905

RESULT 22
SERIABIO/c
LOCUS              4254 bp      DNA      linear      BCT 30-JUN-1993
DEFINITION        S.erythraea erythromycin A biosynthesis gene cluster encoding
                  hydrolase; thioesterase; methylase.
ACCESSION          X60379.1 GI:48941
VERSION            X60379.1 GI:48941
KEYWORDS            erythromycin biosynthesis; hydroxylase; methylase; thioesterase.
SOURCE             Saccharopolyspora erythraea
ORGANISM            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
                  Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.
REFERENCE           1 (bases 1 to 4254)
AUTHORS            Haydock, S.F., Dowson, J.A., Dhillon, N., Roberts, G.A., Cortes, J. and
                  Leadlay, P.F.
TITLE              Cloning and sequence analysis of genes involved in erythromycin
                  biosynthesis in Saccharopolyspora erythraea; sequence similarities
                  between Eryg and a family of S-adenosylmethionine-dependent
                  methyltransferases
JOURNAL            Mol. Gen. Genet. 230 (1-2), 120-128 (1991)
MEDLINE            92079886
PUBMED             1840640
REFERENCE           2 (bases 1 to 4254)
AUTHORS            Haydock, S.F.
TITLE              Direct Submission
JOURNAL            Submitted (19-AUG-1991) S.F. Haydock, University of Cambridge,
                  Tennis Court Road, Cambridge CB2 1QW, UK
FEATURES           source
1..4254
/organism="Saccharopolyspora erythraea"
/mol_type="genomic DNA"
/strain="NRL2338"

```

```

/db_xref="taxon:1836"
69..1313
/gene="orf1"
CDS
69..1313
/gene="orf1"
/codon_start=1
/transl_table=11
/protein_id="CAA42926.1"
/db_xref="GI:48942"
/db_xref="SPTREMBL:Q00463"
/translation="MIFLVGLCKRCIGNHDLSEVLHGDQALTGVPFRNDEIVPSV
PIELVKAPPGGLVQLRESADFGIMYNEGVSRSIGIRPFMNLHGLGVAKLKGWVPV
GPDLDVVDIGSLVLLRGLYLPDAKLAGFDLVGKPRDLYPEADLYTGFSADAFE
EYGERAKVTSIAMFYDLPMPFRMDVHDLDADGLWLMSEQSYLPSMDAGAYDV
VCHLELYGLAQIEWMAQRVGLKVDAEITDVYGGSRVAVLPKQSGHPVDERAGLR
IARFAAKLDTMAYEAFARETERQDQLLEFLAKSRAEGLTLIGYGASTKGNVILQ
YGLTEQDLPFCIGVSPKSGCYTPTGTRIPVSEBAKRRPQQLLVLPFIYRDEFE
RQEFLAGGKKLIFLPLREV"
1310..2530
/gene="orf2"
CDS
1310..2530
/gene="orf2"
/transl_table=11
/codon_start=1
/product="hydroxylase"
/protein_id="CAA42927.1"
/db_xref="GI:48943"
/db_xref="GOA:Q00441"
/db_xref="SWISS-PROT:Q00441"
/translation="MTVPDESDFHVDWVYAELETRAPVTPVRFVFLGDAMLVTG
YDEAAALSDLRLSDPKKVPQVEFPAYLGGFVDVNTFATNMGSDPPTHTLR
KLVSQETVRRVYEMKRPVQETIAELLDEVGSDVVDIVDFAPHLPIKVIKCELLGVD
KYRGFGESEIIVMDPERAEORGQAREVNFILDLVERRTEPGDGLLSALIRV
QDDGRLSADLTSLVALLAGFEASVLIGICTYLLTHPDQQLAVRDPSPAL
PNAVEILRIYAPPTTTPAAEEVEIRGVALPQYSTVLVANGAANRPKFPDPHFR
DVTRGDHLPSFGQGHFCMGRPLAKLEGEVALRFGFRFALSIGIDADDVVRRSIL
LRLGDHLPFLRLDG"
2527..3270
/gene="orf3"
CDS
2527..3270
/gene="orf3"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="CAA42928.1"
/db_xref="GI:48944"
/db_xref="GOA:Q00442"
/db_xref="SPTREMBL:Q00442"
/translation="MSTWLRREGPPVHEEARLVCPFHAGAADSVLDLARALAPIDV
DAVQYEGQRDRDEEPLGTAGEIADEVAALVRASGDGFFALFGHSMGALIAETARR
LEREPGGPLRVFSGQTAPVRHERNTDLPDGLVDELRLGTSEALADEALLAMS
LPVLRADYRLVRSYAWADGPPLRAGITLPGDADPLTATGDAERLQHSVIPGRTTF
PGGHFYLGHQVTEVAGAVRDLIRAGLAG"
complement(3247..4167)
/notes="orf4"
/codon_start=1
/transl_table=11
/product="methylase"
/protein_id="CAA42929.1"
/db_xref="GI:581650"
/db_xref="GOA:Q54095"
/db_xref="SPTREMBL:Q54095"
/translation="MSVKQSKALQDLVDFAKVHWVTRVPPSSARLAVLALPADDHEAT
TEGAINLGVFKPGCAGLEAQNQLANQAEAGASEGDEVLDVDFGLGAQDFWLET
RKPARIVGDLTSPSHVRIASERAERENQDQFKEGSATDLPFGARTFDRVTSLESA
LHNPRTDFPKGAFVLPKPGVLAIGDIIPLDLREFSGDGPPLKAPQSRGSLSGIPV
ENWVPRETYAKLQREAGFVDEVKVDNVMBFWLDYMLRKLQDESFKKSVSRLPYSQ
VKRSLTSDSGMKGLPDLFVIASARKPCA"

gene
Query Match 21.8%; Score 108.2; DB 1; Length 4254;
Best Local Similarity 55.8%; Pred. No. 2.4e-05;
Matches 206; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

ORIGIN

```

```

Qy 130 GGTGAGCTCGTACGTCGCGGGGACGTACAGGATCCACTGTGCGCGAGCCCGCGG 189
|||
Db 1318 GGTGCTCATACGACTTCCAGTCGGGTAGGGGGAAGATCAGCTTCCGCGCCAGCGAGG 1259
|||
Qy 190 AACTCTGCTCCCTTCCGCTGATCTGTCGGCGTGTTCAGCGGAGGAGCGCGTAG 249
|||
Db 1258 AACTCTGCTCGCGTTCGACGAAACCGCTCGCGGTAGATCCAGGGGACACGACGCTGG 1199
|||
Qy 250 TCACACCGCTCGGGCGGTGAACCGGTCGCGGGTGCACACCGGATGTGCTGCGCGGGGCTG 309
|||
Db 1198 TCGGGCGCGCGACTTGGCTTCTCTCCGACACGATCGGATCCGGGTGCGCGGGGTG 1139
|||
Qy 310 AGCGGGCCCTGCTTGGCGCGGTGCTGTCGACACCCAGGAGACGAGTCCGACCGATG 369
|||
Db 1138 TAGCAACCGGACTTCTCCGGGCTGACCTCCCGATGACGAGGCGAGGCTCTCTCGGTGAGC 1079
|||
Qy 370 CGCGAGAAGTTCGTCAGGTCGGGCTCTTCGCGCTGCGCGCTGACGACACACCGCGCTG 429
|||
Db 1078 CGCAGTACTGCAGGATGACGTTGCCCTTGTGCGGGCGCGTACGCGAGGTCAGCTTG 1019
|||
Qy 430 CCTTCGCGCTTGCAGCGATTCAGAGGGCGAGCAGGTCGCTGCGGATGCCCTCGACGCTG 489
|||
Db 1018 CCTTCGCGCGGGGACTTGGCGAGACTCCAGCAGTTGTCGCGCTGCGCGCTCGGTTTCG 959
|||
Qy 490 GCGCGGAAC 498
|||
Db 958 GCGCGGAAC 950
|||

```

```

RESULT 23
RME603644/c
LOCUS Rhizobium meliloti (Sinorhizobium meliloti) 1021 complete pSymB;
DEFINITION segment 3/6.
ACCESSION AL603644 AL591985
VERSION AL603644.1 GI:15140405
KEYWORDS
SOURCE
ORGANISM
Sinorhizobium meliloti (Rhizobium meliloti)
Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
REFERENCE
1
AUTHORS
Finan,T.M., Weidner,S., Wong,K., Buhrmester,J., Chain,P., Gouzy,J.,
Votholter,F.J., Hernandez-Lucas,I., Becker,A., Cowie,A., Gouzy,J.,
Golding,B. and Puhler,A.
The complete sequence of the 1,683-kb pSymB megaplasmid from the
N2-fixing endosymbiont Sinorhizobium meliloti
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9889-9894 (2001)
21396508
JOURNAL
MEDLINE
PUBMED
11481431
REMARK
epub ahead of print
REFERENCE
2 (bases 1 to 315000)
AUTHORS
Weidner,S.
TITLE
Direct Submission
JOURNAL
Submitted (07-JUN-2001) Weidner S., Universitaet Bielefeld,
Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld,
Germany
COMMENT
Submitted on behalf of Universitaet Bielefeld, Biologie IV
(Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and
Department of Biology, McMaster University, 1280 Main Street West,
Hamilton, Ontario, L8S 4K1 Canada
mailto:Stefan.Weidner@Genetik.Uni-Bielefeld.DE
PEXO, pSymB.
FEATURES
Location/Qualifiers
source
1..315000
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSymB"
complement(52..405)
/gene="SMb20542"
gene

```

gene  
CDS  
complement(52..405)  
/gene="SMB20542"  
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions"  
/note="Product confidence : putative  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative protein in ISRM14"  
/protein\_id="CAC48921.1"  
/db\_xref="GI:15140406"  
/db\_xref="GOA:Q9X6A1"  
/db\_xref="SPTREMBL:Q9X6A1"  
/translation="MIVAGQRLPILIAITRPVDRCGHQALAMVOTELKLDPHSGVTYV  
IFRSKGRDLKILYDGTGMVLYTKILEHSPANPKYQDGTMLRSQYEALEPGLDW  
RVNVAQRTPASAAQ"  
289..705  
/gene="SMB20544"  
289..705  
/gene="SMB20544"  
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions"  
/note="Product confidence : putative  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative protein in ISRM14"  
/protein\_id="CAC48922.1"  
/db\_xref="GI:15140407"  
/db\_xref="GOA:Q9X6A0"  
/db\_xref="SPTREMBL:Q9X6A0"  
/translation="MRVELQLQHLHQSLRMPTAEVHRPCNQDRQSLARDHDTGARM  
ARTAAATSTGTGPTRTNTSGPISNVKTDETASDGGAKTFTASLGSAAIAKGTNAG  
SASPLSDRPGIASNGARMPSCRKRQKESWCGTTS"  
complement(402..845)  
/gene="SMB20543"  
complement(402..845)  
/gene="SMB20543"  
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions"  
/note="Product confidence : putative  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative protein in ISRM14"  
/protein\_id="CAC48923.1"  
/db\_xref="GI:15140408"  
/db\_xref="GOA:Q92W24"  
/db\_xref="SPTREMBL:Q92W24"  
/translation="WAGDGFVGRYEVPRGRNRPDPAKARIVASSLEPGYVVDV  
ARRHDVPHQUSFWRRQREGILALPFEMFGLSESGDAEPFVPLAIAEPSEAVNV  
LAPPISEAVSVLTLEIGFDVVVGVFVERVAALVRAMRPV"  
complement(1438..1656)  
/gene="SMB20545"  
complement(1438..1656)  
/gene="SMB20545"  
/function="MISCELLANEOUS; Hypothetical/Partial homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="HYPOTHEITICAL PROTEIN"  
/protein\_id="CAC48924.1"  
/db\_xref="GI:15140409"  
/db\_xref="GOA:Q92W23"  
/db\_xref="SPTREMBL:Q92W23"  
/translation="MDDRAGSAADAEPAPDISEVAQCAAAADLLDFAAQLAGLAEDL  
KALAARPIPTPTLAEPPHPPEATSSDL"  
complement(1843..2361)

gene  
CDS  
complement(1843..2361)  
/gene="SMB20546"  
/function="MISCELLANEOUS; Hypothetical/Global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHEITICAL PROTEIN"  
/protein\_id="CAC48925.1"  
/db\_xref="GI:15140410"  
/db\_xref="GOA:Q92W22"  
/db\_xref="SPTREMBL:Q92W22"  
/translation="MAEKWDECVIIDLENLEGIQIVMSPAYAAQLLENWPDITGRSY  
SAAALNSCTDMFGAASAPARDAFIAAVEARVKTLMHHAAPLVKGNDSIEGGYVA  
ELGFGAARSHSVRANSTGGGRQADREFSARSSRSSESGAVLVSPSTDARLISFTLPPA  
ASRTCRAYSAHW"  
complement(2518..2796)  
/gene="SMB20547"  
complement(2518..2796)  
/gene="SMB20547"  
/function="MISCELLANEOUS; Hypothetical/Partial homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="HYPOTHEITICAL PROTEIN"  
/protein\_id="CAC48926.1"  
/db\_xref="GI:15140411"  
/db\_xref="GOA:Q92W21"  
/db\_xref="SPTREMBL:Q92W21"  
/translation="MYEPFAQIKAWLKEGFTLSAGVLRRLTEIDPIRFTKNIEMV  
QMLVKAWMEMAVLVLDGWTKSLPSAAAAADVTGTSASAVLGNLR"  
3930..4424  
/gene="SMB20548"  
3930..4424  
/gene="SMB20548"  
/function="MISCELLANEOUS; Unknown"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="HYPOTHEITICAL PROTEIN"  
/protein\_id="CAC48927.1"  
/db\_xref="GI:15140412"  
/db\_xref="GOA:Q92W20"  
/db\_xref="SPTREMBL:Q92W20"  
/translation="MESIERNCRKDFDIFMPSFYTELHERTKQILQKRPFPLVGY  
GFVSLPRLNFEELRRVDGVVCFPLRGANGYHGFPGYHGRPVLRSRATAGLPLOAAL  
PEGDRAPRANPALGRPADPAKGSERARLHGRPCRECYRFSKPADQASADNYQWIE  
QPHC"  
6048..6365  
/gene="SMB20549"  
6048..6365  
/gene="SMB20549"  
/function="MISCELLANEOUS; Hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHEITICAL PROTEIN"  
/protein\_id="CAC48928.1"  
/db\_xref="GI:15140413"  
/db\_xref="GOA:Q92W19"  
/db\_xref="SPTREMBL:Q92W19"  
/translation="MRALKASAHYDYMSEQIRALPVQOLASMNCLIVSRATAPHLA  
TVTECLKAWGFYKSPMAWKRTAGKVRNGTGRVTRTTGIVFGVTGLNGPKQSHVFP  
TIS"  
complement(6533..6778)

gene  
CDS  
complement(52..405)  
/gene="SMB20542"  
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions"  
/note="Product confidence : putative  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative protein in ISRM14"  
/protein\_id="CAC48921.1"  
/db\_xref="GI:15140406"  
/db\_xref="GOA:Q9X6A1"  
/db\_xref="SPTREMBL:Q9X6A1"  
/translation="MIVAGQRLPILIAITRPVDRCGHQALAMVOTELKLDPHSGVTYV  
IFRSKGRDLKILYDGTGMVLYTKILEHSPANPKYQDGTMLRSQYEALEPGLDW  
RVNVAQRTPASAAQ"  
289..705  
/gene="SMB20544"  
289..705  
/gene="SMB20544"  
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions"  
/note="Product confidence : putative  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative protein in ISRM14"  
/protein\_id="CAC48922.1"  
/db\_xref="GI:15140407"  
/db\_xref="GOA:Q9X6A0"  
/db\_xref="SPTREMBL:Q9X6A0"  
/translation="MRVELQLQHLHQSLRMPTAEVHRPCNQDRQSLARDHDTGARM  
ARTAAATSTGTGPTRTNTSGPISNVKTDETASDGGAKTFTASLGSAAIAKGTNAG  
SASPLSDRPGIASNGARMPSCRKRQKESWCGTTS"  
complement(402..845)  
/gene="SMB20543"  
complement(402..845)  
/gene="SMB20543"  
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions"  
/note="Product confidence : putative  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative protein in ISRM14"  
/protein\_id="CAC48923.1"  
/db\_xref="GI:15140408"  
/db\_xref="GOA:Q92W24"  
/db\_xref="SPTREMBL:Q92W24"  
/translation="WAGDGFVGRYEVPRGRNRPDPAKARIVASSLEPGYVVDV  
ARRHDVPHQUSFWRRQREGILALPFEMFGLSESGDAEPFVPLAIAEPSEAVNV  
LAPPISEAVSVLTLEIGFDVVVGVFVERVAALVRAMRPV"  
complement(1438..1656)  
/gene="SMB20545"  
complement(1438..1656)  
/gene="SMB20545"  
/function="MISCELLANEOUS; Hypothetical/Partial homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="HYPOTHEITICAL PROTEIN"  
/protein\_id="CAC48924.1"  
/db\_xref="GI:15140409"  
/db\_xref="GOA:Q92W23"  
/db\_xref="SPTREMBL:Q92W23"  
/translation="MDDRAGSAADAEPAPDISEVAQCAAAADLLDFAAQLAGLAEDL  
KALAARPIPTPTLAEPPHPPEATSSDL"  
complement(1843..2361)



```
/gene="SMB20550"
complement(6533..6778)
/gene="SMB20550"
/function="MISCELLANEOUS; Hypothetical/global homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC48929.1"
/db_xref="GI:15140414"
/db_xref="GOA:Q92W18"

Query Match
Best Local Similarity 21.1%; Score 105.6; DB 1; Length 315000;
Matches 195; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 154 GGCAGCTACAGGATCCACTGTCGCGCAGCCCGCGGAATCTCTGCTCTTCCCTCCCATGATC 213
Db 125772 GGCACAGGAATGATAAACTTCCACACCGCTGTGATAGTCGGCTTGTGGCGGAGATC 125713
QY 214 TCGTCGGCTGTTCCAGCGAGACGCGGTAGTCCACCGCTCGCGCGGTGAACGG 273
Db 125712 TCGTCAGCGAAATCCAGGCAAGACAGACGATGTCGGGTGGCGTTGSAAGGGCC 125653
QY 274 TCCGGGGTGCACCGCGGATGTGCTGCGCGGGGTGAGCGGCCCTGCTTGGCGCGCTC 333
Db 125652 TCCGTCGAGCTATCGCGAGAGCTGCGCAGGATATCTTGCCTTGTGGACGTTG 125593
QY 334 GTGTCGCACACCCAGAGACCGAGTCCGACCGATCCGACGATCCGCAAGTTCGTCAGGTGGCG 393
Db 125592 CGGTCGACACGAAAGTCTTAGAGTTCACGTCGATCCGATCCGATCGTGTGATCAGCGTCGT 125533
QY 394 CTCTTCGCGCTCGCGCGCTAGCGCACACCGCTTCCCTCGCGCTTGGCGGAGTTCAGC 453
Db 125532 CCTTGGCGCGCGCCCAAGCGGCAATCGAGGAGCCCTCTGATTCAGCCTGTGCACG 125473
QY 454 AGGCGAGAGGTCGTCGCGGATGCGCTCGAGTCGCGCGCGCAA 497
Db 125472 AGTGCCAGAGCTCAGCTTCAATACGTCGACCTTCGACAGAA 125429

RESULT 24
AF147704
LOCUS
DEFINITION
Streptomyces fradiae NDP-hexose 2,3-enoyl reductase TylCII
(tylCII), NDP-hexose 4-ketoreductase TylCIV (tylCIV), NDP-hexose
3-C-methyltransferase TylCIII (tylCIII), mycarosyl transferase
TylCV (tylCV), and NDP-hexose 3,5- (or5-) epimerase TylCVII
(tylCVII) genes, complete cds.
AF147704
AF147704.1 GI:5305790
Streptomyces fradiae
Streptomyces fradiae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 5908)
Bate.N., Butler,A.R., Smith,I.P. and Cundliffe,E.
The mycarose-biosynthetic genes of Streptomyces fradiae, producer
of tylosin
Microbiology 146 (Pt 1), 139-146 (2000)
20121747
PUBMED
10658660
REFERENCE
2 (bases 1 to 5908)
Bate.N. and Cundliffe,E.
Direct Submission
Submitted (30-APR-1999) Biochemistry, University Of Leicester,
University Road, Leicester LE1 7RH, UK
LOCATION/Qualifiers
1..5908
/organism="Streptomyces fradiae"
```

```
/mol_type="genomic DNA"
/strain="759235"
/db_xref="taxon:1906"
complement(80..1069)
/gene="tylCII"
/notes="orf10*"
complement(80..1069)
/gene="tylCII"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/transl_table=11
/product="NDP-hexose 2,3-enoyl reductase TylCII"
/protein_id="AAD41821.1"
/db_xref="GI:5305791"
/translation="MSGMYVOLGRGATLVSRMLGTWNFGSRVDPDDAIRLMDEALDR
GVNCDVADIRYGLRWLYKGYHTELVGRWLGSRDDVYLAATKVGPEMSDRVNDGLSA
RHVIRCEADIRYGLWYHTELVGRWLGSRDDVYLAATKVGPEMSDRVNDGLSA
GWHLAQCGAARSGSLGSEOCVNLVRAHELEVLPAARAYGIGVFWASPLHGGI
LSGALRLAGSTAVKSCQGAORTLPALRTIARYERFCARVGEDPAEVGLAMLSRP
GVSGAVTGPRTTCHLVSAFRAVELELESEEHRELEALFPFVSGSGVEPEAQN"
complement(1066..2109)
/gene="tylCIV"
/notes="orf9*"
complement(1066..2109)
/gene="tylCIV"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/transl_table=11
/product="NDP-hexose 4-ketoreductase TylCIV"
/protein_id="AAD41822.1"
/db_xref="GI:5305792"
/translation="MSMPLVGPARGRRRLVLLGASGFLGSAVAELARRPVRLRL
VAREPSPVAGAAVELEVRBDLSLPEVTEAVEADAVVHLVAHTGGEKSWAAGER
SEHNVGLMRELAELRPTGPDATRAPVVLFGSTLQAGMEQATPGTYAAQKLAER
HLRADAGRVGVLLRTTVIGRSPITGSPRGVIAVTAAGALAGDPIITWHDGSEV
RDLDDVDAATAFTTALHADQLGKHVVGTGRRHRLDRVFGTVAALAAHGTGRPEV
PVTVDPGVAEFCDFRTPDPSAFRAVNTGMRPRAEPADGIAAATAAAGAGDSPEP
EGAGKGGGTER"
complement(2106..3362)
/gene="tylCIII"
/notes="orf8*"
complement(2106..3362)
/gene="tylCIII"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/transl_table=11
/product="NDP-hexose 3-C-methyltransferase TylCIII"
/protein_id="AAD41823.1"
/db_xref="GI:5305793"
/translation="MPAVPRDQMIISACVCGNRELLPVLDELGEQALTGVPFRTREE
TVPSIPLLVKSPQCGVLQRLRTPDGLMYGEGYGRSGIRPFMHLRRVAAIR
ELVDLGPDDLVDIGSNDATLKAYPEGRLVIGIDPSGDKRFLYPPHAEILAYFSR
DVTAREGTRARVITSIAMFYDLPDPLAFMRDVHDVLDAGDGMWEOQSLPAMLEAD
AYLVCHHELYVALQCIEMAEAGLTULRSLTYVYGGSLCVTLARASSPHRDEA
GPARIARETAKINTMAPFEFARVHORDALRDFLDKRAARGRLTLCGASTKGN
VILQYCGIGERDLPCIGVSPERAGRTTGTGTGIPVSEEDAKAMPDQLLVLPWYRE
GFVEREDFLAGGRLVFPPLRLDV"
3575..4960
/gene="tylCV"
/notes="orf7*"
3575..4960
/gene="tylCV"
/function="catalyses the incorporation of mycarose during
tylosin production"
/notes="glycosyltransferase"
/codon_start=1
/transl_table=11
/product="mycarosyl transferase TylCV"
/protein_id="AAD41824.1"
/db_xref="GI:5305794"
```

```

/translating="MAGLRPGAGVPPGTPWPIPSGKHGCVRSATSRRLRPLYPHAGTPL
GRERRADEGSCVVAHTAFILPAAGHVAHTPLGVAELAAHGHVTVVLPEDMDARAV
RVGARVYPLDRFRADNVKPESEDEYDEGEFLKYLEWLLDTADTTLLESAFA
EDPDVANDPSTFWKLLAGKWDIPVIRKSTFASNEHWAHPPEFGCAQVDPAL
IETARAEKLLKHEGTTSDVAFAATVQSGPLFYMPRYFQVAGETDRHHRFVGSCA
PRASFHTWQRPEDRGLVMSLGTINERPGFACVAFDRPWNILFLVGGHVGGA
GDAGPLPENLVDRFVPLGVDLPHDTHLLVNHGCTAMEALAHGVPVIVAMPPEPRA
TARIABLDLGDWLLPGEVTAELKSLGIAQRLVLTDDIRIKGLDRMRGEIRRAGGPAVA
DVIEGLLSPPAA"
5004. 5597
/gene="tylcvii"
/notes="orf6"
5004. 5597
/gene="tylcvii"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/translating="NDP-hexose 3,5- (or5-) epimerase Tylcvii"
/protein_id="AAD41825.1"
/db_xref="GI:5305795"
/translating="MIITETVRDAVRAITPEIPDRHGSILYSLRYETLRRATGHAIE
IQVHNVNRRNTLGRHGTVPVPGQKIVTCVRAARTWVLDLRVGSPTFGHGVVG
DQAGVAVVLPDLGLGYVALADTGMNLYLTREYTPGMIIDIDALDPLGLPWNLT
EPVRSERDAAPSLAEAAAGTLPGEQCLRAYPA"
ORIGIN
Query Match 20.8%; Score 104.2; DB 1; Length 5908;
Best Local Similarity 55.7%; Pred. No. 7.1e-05;
Matches 199; Conservative 0; Mismatches 158; Indels 0; Gaps 0;
142 ACCTGTACCGCGGGGACGTACAGATCCACTGTCGCGGAGCGCGGCGGAACTCTGCTCC 201
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
2113 ACCTGACCGCGGGGACGGGGAACACAGTCCGCGCGCGCGCGGCGGAGAGTCCCGTCC 2172
Qy |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
202 TTCTGCGCATGATCTGTCGCGGTGTTTCCAGCGAAGACAGCGCGTGTAGTCCACCGCTGC 261
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
2173 CGCTGACGAAACCTCCCGGTAGATCCAGGCGAGGACGAGTGTGTCGCGCGCATC 2232
Qy |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
262 GCGGTGAACGCTCGCGGTGCGACCGGATGTCGTCGCGGGGTGACCGCGCCCTGC 321
Qy |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
2233 GCCTTGGCGTCTCTCTCGAGACGATCGGATGTCGCGGAGTACCGGGGGTGAACCGCGCCCT 2292
Qy |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
322 TTGGCGCGCGTGTCTGCCACACCCAGAGACAGGTCGCGGACCGATGCCGCGAGAGTTC 381
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
2293 TTCTCGGGGTCACCTCCCGATGACAGGAGTCCCGTTCGCGATCCCGCAGTACTGG 2352
Qy |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
382 GTCAGCGTGGCGCTCTTCCGCGTCCGCGGTACGCCACCGCGTTCCTTCGCGCTTG 441
Qy |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
2353 AGGATGACATTGCCCTTGTGAGCGCGCGTACGCGAGAGTACGCGCGCGCGCGCGG 2412
Qy |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
442 AGCGAGTTTCAGCAGCGCGGACGAGTCCGTCGCGATGCCCTCGAGTCCGCGCGGAAC 498
Db |TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|TTT|
2413 GACCGGTCAGGAAGTCCGCGAGGCGTCCGCTGCTGACCCCGCGGCGGAC 2469

```

```

RESULT 25
AE008206/c 12862 bp DNA linear BCT 18-DEC-2001
LOCUS Agrobacterium tumefaciens str. C58
DEFINITION 187 of the complete sequence.
ACCESSION AE008206
VERSION AE008206.1
KEYWORDS GI:15158374
ORGANISM Agrobacterium tumefaciens str. C58 (Cereon)
SOURCE Agrobacterium tumefaciens str. C58 (Cereon)
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
1 (bases 1 to 12862)
REFERENCE Hinkle, G., Slater, S.C. and Goodner, B.
AUTHORS Complete Genome Sequence of Agrobacterium tumefaciens C58
TITLE (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 12862)
AUTHORS Hinkle, G., Slater, S.C. and Goodner, B.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
COMMENT Street, Cambridge, MA 02139, USA
COMMENT Approximately 800 bp of telomeric sequence missing from the left
COMMENT end of the chromosome and 200 bp missing from the right end.
FEATURES
Location/Qualifiers
1..12862
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/mol_type="genomic DNA"
/strain="C58"
/db_xref="taxon:181661"
174..1466
/gene="AGR_L_168"
174..1466
/gene="AGR_L_168"
/notes="hypothetical protein slr1610 - Synechocystis sp.
(strain PCC 6803)"
/codon_start=1
/translating="1"
/product="AGR_L_168p"
/protein_id="AAK88650.1"
/db_xref="GI:15158375"
/translating="WNEAVFSARLSTQHVGLCLCLNSPLVHTFVLDGMSPPCESFV
RADAANDVEPPYPLHAFYCDCEFLVQVQVAPENIPEYAFSPSDSWAHAKRYC
DMVIERFALGQSSFFVLEASNDGILLQHLFSLNIPMLGIEPAVNAVAKVAIGKIEPLT
EFFNEALATDAAACQKADLIIGNNVLAQVDINDFVAGMKTLKPKGVITIEFPHIE
KLIENQPDITTYHEFSYSLTTERMARHGLKVFVBEIPTHGGSURVFPFSDIR
FPREARVDLARELNAGLKIETVYFAEAVRQTKRNLFLRLKEMRKSICAYGA
PQKNTILNYCIGITDIDFAVDNRPYKHGRLTPGMHPIRPVSEIPIPMKDDYVLILP
WLNKEIYVAQMKDIRNKGKFTVPIPDISIIDPKELVQ"
1463..2275
/gene="AGR_L_170"
1463..2275
/gene="AGR_L_170"
/notes="putative transferase"
/codon_start=1
/translating="1"
/product="AGR_L_170p"
/protein_id="AAK88651.1"
/db_xref="GI:15158376"
/translating="MKVVLFCGGRGTRIREYSESVKPLIPLGSPINRHWQYARY
GHEDFILCLGYKAVHKEFFLNSRPESYADCVISSPGRVLELLEPRMVRSLIDTGI
WNIIGERLWAVRHHLKGMFLYNSDGLTDVLDMDIAQFESKDLACFLAVRPLTI
YHLAIDBSGKVRBFSSSETSDMWINGVFLMRPEIPLYMREGBELVLEPFRSLISE
KLWAYKHGSPWRSMDTLRDWQTLFAMVERGDMPIWGDDASMARKSSNAKLIP"
2215..2925
/gene="AGR_L_171"
2215..2925
/gene="AGR_L_171"
/notes="hypothetical protein SCP62.20"
/codon_start=1
/translating="1"
/product="AGR_L_171p"
/protein_id="AAK88652.1"
/db_xref="GI:15158377"
/translating="MDWRCVNGTQIVKQAAAMNMSLASAGDRSLISLCLGAHSDDI
EIGVGGTIVLSLKGVLIDVYVWLSAGNERALEARSAAFLAKAASYSIELASFQD
SYVPSOSRIKSWLLTRDKVSPDIIFTHSREDHQHREINRLTSNIFNLHILEYE
IPKWDGFRGRNAYVPLSEDAAMMKVELLLQHFQTSQSKMDPDAFTFGLRLRGMEC
BAPTRPAEAPTVRKLTPA"
2865..3722
/gene="AGR_L_173"
2865..3722
/gene="AGR_L_173"
/notes="putative response regulator"
/codon_start=1
/translating="1"
/product="AGR_L_173p"
/protein_id="AAK88653.1"
/db_xref="GI:15158378"

```



[illegible]

PDIIPTHSHDAHDHREINRLTSGNPRNHLILVEYIPKMDGDFGRNNAVPLPSDAM  
NMKVELLQHEGTSKDMFADATFRGLARLRGMECAPTEFAFTVRKLTPA"  
complement(7499..8311)  
/gene="rfbf"  
/note="synonym: Atu4797"  
complement(7499..8311)  
/gene="rfbf"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="glucose-1-phosphate cytidyltransferase"  
/protein\_id="AAL45591.1"  
/db\_xref="GI:17743309"  
/translation="MKVILFCGGRGTRIRSEYSEVPKPLIGSQPIMRWQYIARY  
GHEDFILGLYKAVHILKEFFLNSPESVADCVISPGRVLELLELPDRWVSLIDTGI  
WNIIGERLWVPHILKGEKFLANYSDGLTDVLDLDMIAOFKSKLACFLAVRPPLT  
YHLAIDSGKVRPERSSETSDMNINGYFLMRPEIDYMRGELVLEPRLISED  
KLMAKSGFWRSMOTLRDQWLEAVRGMVWIGDDASNAKSLNP"  
complement(8308..9600)  
/gene="Atu4798"  
complement(8308..9600)  
/gene="Atu4798"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAL45592.1"  
/db\_xref="GI:17743310"  
/translation="WNEAVPSARLSQHYGLCLNSPLVHTFVDLGMSPCESFV  
ADAANDVEFPYPLHFAVDECEFLVQLQEYVAPENIFTEYAFSFSDSVYAKRVC  
DMVIERFALPQSVFVELASNDGLYLLHFLTNSIPMLGIEFVAVNAKVAIGKIPTLT  
EFNEFALATDMAARKADLIIGNVLAQVINDFVAGMKTLKPEGVITLEPHIE  
KLIAEQDITTHYFYSLSLTIEKMARHGLKVFDEETPTGSGSRVTFSEHDIR  
PFEAREVSLAELNAGLKIETFAFAEAVRQTKRNLFLRLKEMKRSICAYGA  
PGKGNLNYCGIGTDFIDFADVRNPYKHRLTPGMHIPRVSIEIPHKMDYVLLP  
WLNKEIWAQMDIRWCKGFVPIPIIDSKELVQ"  
9305..10432  
/gene="rfbc"  
/note="synonym: Atu4799"  
9305..10432  
/gene="rfbc"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="dUDP-4-dehydroxamnose 3,5-epimerase"  
/protein\_id="AAL45593.1"  
/db\_xref="GI:17743311"  
/translation="MLFEKTKIEGLLVNTIPVDRSRGSFARTFCKREFESNALVSDF  
SQHLSLSKLKHTIRGLHPQPHQETKLVSCVKGSINDVAVDLRRGSPTYLQWGTV

Query Match 19.6%; Score 98.2; DB 1; Length 12952;  
Best Local Similarity 54.6%; Pred. No. 0.00034;  
Matches 196; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 140 TCACGTGTACCGGGGACGACAGGATCCACGTCCGCGAGCCCGCGGAACTCTGCT 199  
Db 8334 TGATGATATGTCGGGATTTGGAACGATGATTTCCACCCAGTTTCATGATCTTTCA 8393  
QY 200 CTTTCGCCATGATCTGTCGGCGTGTTTCAGCGAAGAGACGCGGTAGTCCACCGGT 259  
Db 8394 TTTTCGCCACGATCTATTCTTCAGGTTCCAGGGGAGATGAGAACGTAGTCCGGCTTCA 8453  
QY 260 CGGCGGTGAACGGTCCGGGGTGGCACCGGGATGTCGTGCGGGGGTGAGCCGCCCT 319  
Db 8454 TGTGCGGATTTCTGACACGGCGGATCGGGATATGCAATCCGCGGTGAGACGACAT 8513  
QY 320 GCTTGGCCCGGTGGTGTGCAACACCCAGAGACCGAGTCCGACCGAGTCCGAGAGT 379  
Db 8514 GTTTGTAAAGGTTTCGATCGACTCGCAAAATCGATGAATCGGTACCTATCGCCCAATAGT 8573

QY 380 TCCTCACGGTGGCGCTCTTTCGCGGTGCGCCGTAAGCCACACCGCTTTCGCTCGGCT 439  
Db 8574 TCAGCAGCGTGTTCCTTCCCGGTGCGCCATAGCGCAGATCGATTTGGCATCTCT 8633  
QY 440 TGAGCGAGTTCAGCAGCGGAGCAGTTCGTCGGGATGCGCTCGAGTTCGGGCGGAC 498  
Db 8634 TGAGCGGATGAGGAGGACAGGAGTTTCGCTTGGTTCGCGCACGGCTTCGCGGAC 8692

RESULT 27  
SCO939105/c  
LOCUS SCO939105 291000 bp DNA linear BCT 11-FEB-2003  
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 2/29.  
ACCESSION AL939105 AL109374 AL117385 AL117387 AL117669 AL121596 AL121600  
VERSION AL121719 AL121849 AL121855 AL122662 AL132707 AL645882  
KEYWORDS AL939105.1 GI:24418991  
SOURCE Streptomyces coelicolor A3(2)  
ORGANISM Streptomyces coelicolor A3(2)  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,  
Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,  
Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,  
Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,  
Huang, C.H., Kieser, F., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,  
Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S.,  
Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, J.,  
Taylor, K., Watters, T., Wietzorrek, A., Woodward, J., Barrall, B.G.,  
Parkhill, J. and Hopwood, D.A.  
TITLE Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)  
JOURNAL Nature 417 (6885), 141-147 (2002)  
MEDLINE 21996410  
PUBMED 12009953  
REFERENCE 2 (bases 1 to 291000)  
AUTHORS Bentley, S.D.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces  
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
COMMENT On or before Oct 29, 2002 this sequence version replaced  
GI:20520891, GI:20520894, GI:20520895, GI:20520772, GI:20520896,  
GI:20520897, GI:20520883, GI:20520899, GI:20520902.  
FEATURES  
source  
1..291000  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
213..3092  
/gene="SCO0305"  
/note="synonym: SC5G9.14"  
213..3092  
/gene="SCO0305"  
/note="SC5G9.14, hypothetical protein, len: 959 aa;  
unknown function, similar to TR:O54182 (EMBL:AL021411)  
Streptomyces coelicolor hypothetical protein (913 aa),  
fasta scores; opt: 698 z-score: 776.6 E(): 0, 27.0%  
identity in 972 aa overlap. The C-terminus shows very weak  
similarity to eukaryotic beta-transducins e.g.  
SW:GBB4MOUSE (EMBL:S86124), gnb4, Mus musculus guanine  
nucleotide-binding protein beta subunit 4 (340 aa) (23.0%  
identity in 291 aa overlap). Contains possible hydrophobic  
membrane spanning region"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CA355662.1"  
/db\_xref="GI:5881863"  
/db\_xref="GOA:Q9RL12"  
/db\_xref="SPTREMBL:Q9RL12"

```

/translation="MSDAVEAEAEARAVPVGAKORDAFISYQSQKNAPLAKQLQRLGLE
RVASFRPMWSVFDVTSIPANHLMQSIQRELARGRYFILLASPEAAASPWVARE
VYWLRETRTEHLIAVSGEIVWDRAGDFWRKTSIPRLNLAGHTAAPLWVLDRE
IGERSHSMRYPPRAAVATLASPLHGEPEKELEHDDPEQLFVKRLSWSGVALVAL
LVAVTGFDSRQREAAALSQARTSASQAARSQQLGTSNQAAQYALVAEETEST
PSRRAALAQTAAPFARRRLRADDSVTDMDGSGRPAFADIVLSGDTTAAYSVFD
DRAVALYDVSAGQSVVRABEPTPALSRDGRVRLAMEVILNRVQMDTRSGRLLSUTM
ATENTENLPHVGHSHALALSPDGRWAAASHYTPGEAFVAVWRASDGREITRVPVS
ARVLGFSADGSLTLVDQDRREVRFTDSARWEAARSLPGMPSGTAEALRYSEVLL
FDGAKALTRTSRAEAWDLEQRRIAVRSLGFEYEDMTVADDKGIPLGGQDGTIRLY
DALRPGTIGRLTRPASTAMSDGTQVVAASYGCELNLFTTVGRGQOAVDEKDFV
TFGLDIPDEMAVHREGEETEDPATGRGLGVIPYANLSVLEDTAVTLSDGKRYVQ
MQSVFEGDEGRDWNDRIGRIDCTCTAVGSSSPRIRPGLSILPGLDRIYVGEWG
AVELLTRICTVETSDSGRFALSGDRRLVILNDGDEVRATGLPKDTDAVAFSDGR
PSAHPVMSVDHEGQAAPFDSHVVYVRLDSGRVATGLPKDTDAVAFSDGR
LLPQGFMTSTRTVLDAGSGDQDITWTDPTGVSQVNEGMVQVPGPADDILTLP
DHKVVRTVGVGAHDLICLVLSQPLPARSDRYLKDLDVEDPCR"
/complement(3056..3961)
/gene="SC00306"
/notes="synonym: SC5G9.15c"
/pseudo
/complement(3056..3961)
/gene="SC00306"
/notes="SC5G9.15c, possible pseudogene, len: 302 aa;
similar to hypothetical proteins e.g. SW:YCYJ_ECOLI1
(EMBL:AE00230); YCYJ, Escherichia coli hypothetical
protein (310 aa). GC frameplot and codon usage plots
indicate a coding region which lacks a start codon"
/pseudo
/codon_start=1
/transl_table=11
/product="putative pseudogene"
/length=4971
/protein_id="CAB55664.1"
/gene="SC00307"
/notes="synonym: SC5G9.16"
/length=4971
/gene="SC00307"
/notes="SC5G9.16, possible DNA-binding protein, len: 295
aa; probable CDS suggested by GC frameplot and positional
base preference. Similar to possible DNA-binding proteins
from S. coelicolor e.g. SCJ9A.15c (cosmid SCJ9A) / S.
coelicolor possible DNA-binding protein (290 aa) (39.3%
identity in 298 aa overlap). Contains probable
helix-turn-helix motif at residues 35-56 (Score 1197,
+3.26 SD)"
/codon_start=1
/transl_table=11
/product="putative DNA-binding protein"
/length=4971
/protein_id="CAB55664.1"
/db_xref="GI:5881864"
/db_xref="GOA:Q9RL11"
/db_xref="SPTREMBL:Q9RL11"
/translation="MSSELGDFLKARRELSPPTAGLPADGRRVFGLRREEVAVLAS
ISPGHYTRLEQGRRAASPVLDALARLRSADERAYLFSLGKDGRLGRPGQVR
PQLRELLDLFTHTALVLRCTDILANPPAAALFTDFALLSRDNRNVLVPRDPAV
RALYTDWFMARCAVQRMEARDPDRLSALVLAELSAQDPDPTWTKDHRVAVRG
AGTKDLRHPVIGELFLDWSALTDTADPQQLIALTAAPGTPSHGLRALRDLGAVTSG
RTGPGVGGGGEIAGASQA"
/length=5026
/notes="SC00308"
/gene="SC00308"
/notes="synonym: SC5G9.17"
/length=5026
/gene="SC00308"
/notes="SC5G9.17, possible membrane protein, len: 467 aa;
unknown function, probable CDS suggested by GC frameplot,
positional base preference and amino acid composition.
Contains hydrophobic, possible membrane-spanning regions
near the C-terminus. Contains TTA (leucine) codon;
possible target for bida regulation"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/length=5026
/protein_id="CAB55665.1"

```

```

/db_xref="GI:5881865"
/db_xref="SPTREMBL:Q9RL10"
/translation="MPRSDAARVTRLLMYHLTPQVHVDPDRRLPLSEALGLGRTPT
ARVEAAGLLPRLVHERVREITLPNNAARVDGADITLAVTPQDPLFLVDVEVPEA
GDSPTTAETLYAQWQSRAIRVGGVPLTAWLDERLAYGLTDRPPLPFAFNGVHQ
SVFPGAGLRELLAAHVPDAGVDVITVIRGTLDGSGTATGIRPETINNAGQTM
VAMGRGSLVGVYAPPVENAFGAAGAVTIRGAHRIRQAAALRELDRAAAVETTD
AARALVAQLSHRLNDQLDLSFSEVYVADTILVPELLVSEFASLRVDSLAALANT
SRIVERNVAVIATRVGLETAQAQYADRDRLVLAAVIAGSLALPALLAVFGVNS
SDVPSISILDFQHYGAYIAWSPFLLILLGLLARISLRLPAPDAATSERTRER
FRSLRSLWGRSRPGR"
/length=5434
/gene="SC00308"
/notes="TTA (leucine) codon; possible target for bida
regulation"
/complement(6537..7625)
/gene="SC00309"
/notes="synonym: SC5G9.18c"
/complement(6537..7625)
/gene="SC00309"
/notes="SC5G9.18c, probable amide hydrolase, len: 362 aa;
similar to 6-aminohexanoate-dimer hydrolases which degrade
N-(6-aminohexanoyl)-6-aminohexanoate e.g. PR:P75007
(EMBL:X02336), NylB, Flavobacterium sp. (strain XL72)
6-aminohexanoic acid linear oligomer hydrolase (E11) (392
aa), fasta scores; opt: 531 z-score: 598.0 E(): 6.1e-26,
34.1% identity in 323 aa overlap"
/codon_start=1
/transl_table=11
/product="putative amide hydrolase"
/length=5434
/protein_id="CAB55666.1"
/db_xref="GI:5881866"
/db_xref="GOA:Q9RL09"
/db_xref="SPTREMBL:Q9RL09"
/translation="WPSRRVWRGCGDLPSPGALDGLGVGTGRTTLTDLQLLA
EARTDALVLRHGLVHEEYRHYEYHVPFNASAKSYLGLAATLAHEGLDRGAR
VARYPELSTGAFGDAQVDHLMHGTRMSYAGRPYDKALEQRYFAVLALRFPYGS
GATIREHLATARGEGTEFRYENGVEALAEVLREITGLTSALMSEMIWSRIGA
BEDAYLDAGAEACAGGFSATARDVARIGELRCGAGVDRQVLPAVTPARSVP
DCPRVRFVPAAPDASPATLSYHDLMTLNDPYGFSMASGHGRLFVSPALDLVVH
YASQVSPAVPQVPLVQAPLSIGAHNA"
/length=7816
/gene="SC00310"
/notes="synonym: SC5G9.19c"
/complement(7816..8466)
/gene="SC00310"
/notes="SC5G9.19c, possible tetR-family transcriptional
regulator, len: 141 aa; shows similarity to the N-terminal
half of TR:Q24741 (EMBL:AB001683), FarA, Streptomyces sp.
strain FM-5 receptor for an extracellular regulatory
factor (FM-2) (221 aa), fasta scores; opt: 154 z-score:
193.4 E(): 0.0021, 36.7% identity in 79 aa overlap and to
putative transcriptional regulators. Contains Pfam match
to entry PF00440 tetR, Bacterial regulatory proteins, tetR
family"
/codon_start=1
/transl_table=11
/product="putative tetR-family transcriptional regulator"
/length=7816
/protein_id="CAB55667.1"
/db_xref="GI:5881867"
/db_xref="GOA:Q9RL08"
/db_xref="SPTREMBL:Q9RL08"
/translation="MDMKFEDRRVRTRRAALRKALVELVDKGFQAVTVEITERADI
GRATFYAHYRDKEDILLVGLVRDLAQDRLELPAVQAQAQGTGLPKYIFPHAQER
GVYQVILREGDGRALREFTLIRSHAEMAFDRARQLGVSPRIPIGIVARWTGELI"

```

Query Match 19.4%; Score 97.2; DB 1; Length 291000;  
Best Local Similarity 54.5%; Pred. No. 0.00019;  
Matches 195; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 141 CACGTCAGTCGCGGGGAGCGTACAGGATCCACTCTCCGCCAGCCGCGGAGACTCTCTGCTC 200  
DB 115464 CACGCTCAGTTCGGGTATGGGAGACCGAGCGCGCCGCCACTCTGTCGACGAGGAGAG 115405

```
QY 201 CTTCCGCCATGATCTCGCGGTGGTTTCAGCGGAGAGCAGCGCGTAGCTACACCGCGTC 260
Db 115404 CTTCTCGACCAAGTCTCTCCGCGAGGTTTCACGCGACGAGGACCTAGTCTCGCGCGGTC 115345
QY 261 GGCCTGTGAAGCGTCTCGCGGTGGCGACCGGAGTGTGGTCCGCGGGGTGAGCGCGCCGTG 320
Db 115344 GCGCGGATCCGCTCGCGGTGGCGACGAGTGTGGAATCGCGGTGCGCGGGGTGAACCTGCGGTG 115285
QY 321 CTTGGCGCGGTGCTGTGCGCACACCGGAGACCGAGTGTGCGGACCGATCCGCGAGGTT 380
Db 115284 CTTGTAGGAGTTCGCGGTGCGACCTGTATCGGAGCAGGTGTGGCGCGACCGCGAGTGGTT 115225
QY 381 CTTCCAGGTGGCGCTCTTCGCCGTGCGCGGTGCGGACCGCGTGTCCCTCGCGCCTT 440
Db 115224 GAGCAGCGTGTTCGCTTCGCCGGGCGCGGTATCGGACCGACCGCGTGTCCGCGCGTCCGCG 115165
QY 441 GAGCGAGTTCAGCAGCGCGGAGGTCGCGGTGCGGATGCCCTGACGTCGCGCGGCGAAC 498
Db 115164 CGCGTCGATCAGGACCGCAGCAGGTCCTCCGCGCACCTTGGCCACCGCGGACGAGAAC 115107

RESULT 28
LOCUS AY048670
DEFINITION Streptomyces globisporus enediylne antitumor antibiotic C-1027
ACCESSION AY048670
VERSION 1
KEYWORDS 85163 bp DNA linear BCT 05-NOV-2002
SOURCE Streptomyces globisporus
ORGANISM Streptomyces globisporus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
1. (bases 1 to 85163)
Liu, W., Christenson, S.D., Standage, S. and Shen, B.
TITLE Biosynthesis of the enediylne antitumor antibiotic C-1027
JOURNAL Science 297 (5584), 1170-1173 (2002)
MEDLINE 2171413
PUBMED 12183628
REFERENCE 2 (bases 1 to 85163)
AUTHORS Shen, B., Christenson, S.D., Liu, W. and Standage, S.
DIRECT SUBMISSION
TITLE Submitted (27-JUL-2001) Chemistry, University of California, Davis,
One Shields Ave, Davis, CA 95616, USA
FEATURES
Location/Qualifiers
1..85163
/organism="Streptomyces globisporus"
/mol_type="genomic DNA"
/strain="C-1027"
/db_xref="taxon:1908"
complement(8..658)
/notes="ORF(-7)"
/codon_start=1
/transl_table=11
/product="hydroxylase-like protein"
/protein_id="AA06648.1"
/db_xref="GI:24575077"
/translations="MGWGTGARPVLVVVGAGTATAIRLLHFAHPLEWLLERRAA
YRAGVAYHRDGNPDHVFNIQAGRSVPRVLDYDFINWAOEADRDWPFRWAKK
TEGPAPRRIFODYLDARLVARESCPGVVEADGEALDARSHDRCEFTVRLTTP
YLEGLRPGPLPOTQILDADHVVLATGLELKEPPFAAGVAGHPSFVRNYSAGFIL"
complement(930..1478)
/notes="ORF(-6)"
/codon_start=1
/transl_table=11
/product="viral infectivity potentiator-like protein"
/protein_id="AA06649.1"
/db_xref="GI:24575078"
/translations="MERIPVHTTIDTAPAGGDTILRLERKFRFRLVNIHGMASHPVVL
ETVAITGVAHEGTDFARTREIALAVGADVACQAAHTVSXAVAGTLEETVAI
RRGTGDDVKLEALVQVAREIAGEVGEASDASWNAVAQGWTDTELAIEVHVAVNLY
TNYFNHVARTEIDVPGVPDIDS"

complement(1649..2713)
/notes="Orf(-5)"
/codon_start=1
/transl_table=11
/product="truncated methionine synthase-like protein"
/protein_id="AA06650.1"
/db_xref="GI:24575079"
/translations="MTIATEPIGSLPRSATLLHALAAHQGTLDTATLAKQEQAVAD
TLRLERGVSVLVGQSQSFATYPTVGLDITLSPDGAVIPADGHTROLPSITQGP
FRYQVRAETVLAARQLTDRPLQAVIAPALSILLYPATIEGVPRQFRLDLADEAE
ADRLCDAGAHVQLDFTGRLSKLDPSGVLDLDDFIALNNVGLGFSAEETARIGV
HTCGQDQSTHSLIDYABELPKLFQKAGNFYLEAGADPERVLSIVRDHLPPAA
RVFLGVTDPIPVVETPEQVRDRILLARIYIPVQLGTCDDCGSPFADDTSTTRDLA
PAKTEARVGTALAEVIGL"
complement(2851..3237)
/notes="ORF(-4); similar to SCP1"
/codon_start=1
/transl_table=11
/product="transcription factor"
/protein_id="AA06651.1"
/db_xref="GI:24575080"
/translations="MSSLATDSLVHEAKVWVLEDPGLDYVRQALDKTPRRKNKPRY
ARDGMIGYIELGADAEADPSGLYRRRVFLLPHDRSDSPEGVYRGAFGEAVDPRT
IEPNRVGEKTPRSQLTGSTVATGS"
complement(3442..4971)
/notes="ORF(-3); similar to SCP1"
/codon_start=1
/transl_table=11
/product="putative primase"
/protein_id="AA06652.1"
/db_xref="GI:24575081"
/translations="MSSSARGPRFDATAAQAQMLAETTPAPALLPVQAAPEAPVAA
TAWAGIIPASLTDRGNAKLFLVLYRQFHEVGLGKYSWVGKMGKAGGKAAALWAA
GEMAEPMGSDPGLFTDRHLHKKRLITSTGKALLTQAKASPDLSLDPDITLDDGP
YALCTPGVDLNRMRKPDPTDFHSRATSAPQDIPTRWHRFLEDITPGSDEGR
EMIDFLLHLLYSITGDVGAQVLPFLHGQKNGSVLLDVMIIQILGDAAPPGFLM
DRGAYSEHSTELTGLHRLITVCSLKNDFRDEARVLLTGDKIKARRNRQYFSP
TPTKLWLLGNHRPEVSTGGAFFWRIRLLPFRIVPDERKIDNLVALVDEQEGFIL
HMLTEGARVILATRDITLADPRVLIATSAVANTEDHIGRFLAECCLDHPENSELVEQ
ELLVTSYSTWCANSEGIPTGATRAFAFRVQEVGLASPADMIKSNRKFYFNLAALD
E"
5982..7478
/notes="ORF(-2)"
/codon_start=1
/transl_table=11
/product="glycerol phosphate ABC transporter"
/protein_id="AA06653.1"
/db_xref="GI:24575082"
/translations="MRALPGSGLRMAGDAPTAAGNRCCPTPIQRFLLLRQTLQAP
EAKNGSPQEQRIVLDVNLANYPTDYDRDFLAPLRERAAEFRAHPRYRVLDIN
GHDFWTIPKVARATAEGRPHIATYATDSQLARDARRPDGKVPFTSVEAALAGRT
ILGHPVVLDLPVVRDSYSGBELVSLPTVTMLCYANSLLIARAGVPELPRTWDE
VEAACAQVAVDGGPGHGTWANDGWVQQAVALQNGVLTQDNGRSGSATVVDVTS
EMLDVWRVTHLHSHGHVLTGSPDKAGGAFAVQOKVAFPTDSSKAARELIQAGAQ
AGFEVAFPLPRNKAQVAGQVSGDSIMLAAGLDETQGLLALTQYLLISPANAQM
HRTNGFVPTGAAGELLEATGWFDRFQQRVAGQLKASDKSPAAUGALLGDFAAVNE
VITAMDDVLRSGADPAKAFABAGVAAQQLLDAYNARNRSGSGTSPAS"
complement(7573..9927)
/notes="ORF(-1)"
/codon_start=1
/transl_table=11
/product="uvrA-like drug resistance pump"
/protein_id="AA06654.1"
/db_xref="GI:24575083"
/translations="MYGQENSVSDTPTAGAKTSGPGGIRIAGARHNLKDVSL
TIPRNIQITVTVSGSGKSIPTDVAEQRLNSTFSWYINQLPKYERQAEALIE
NLTPVTVDOKPVGNARSTVGTMDVPMIRALFARFGTTKDDPSLSGVSFSAFND
PQGCPCDGLGQSIALDLMKDRSLDDGAVLFPYKVGSPDWIWKSGSLDPA
KPIAYSAEELDTLLRGTKGKYLTKSTPOTNYEGLARFELIKRDLALSDDRK
REVERFVTDGVCRCRARNAAALDVRIDGKNIADYSNEVRDLTEVLAGVTEPAA
APLAKARTALRIVISIGLYLTIDRPTADLSGEGGRLKMRVHLGSLAGLTIIFDR
PSIGLHPDRGLRDLRLARDKGTVLVWHDPTVIEADIHVIVDVGFGAGVHGEIV
PEGSGFAKLKAKTRTGALRRGARVKEVRPPTGELTVENADLHNLKQSVAVPTGVL
```

TAVTGVAGSGKSLISGAFMEAHDPDAVFVQSOAIAASSRTPVSVYGLMDPLKPLPAK  
BTGNASLFSNKSCEBQCGGVIIITELAFMDPVTHCGVCEGRFRFPVLEHQLR  
GKSTADVLEPADVAFPFTEKALLPKRALVDVGLGYSLGQPLSLSSGGEQLRIKL  
ADQLRGTVTVLDEPTGLHMSDVDTLLKLLDGLVEAGNTVVVHNLNDVVQADWI  
IDLPGDREGREIVFTGTPKDLAADSLSLTGYLRHLKAAQAG"  
complement (9982. .11349)  
/note="ORF0"  
/codon\_start=1  
/transl\_table=1  
/product="Na/H efflux pump"  
/protein\_id="AA06655.1"  
/db\_xref="GI:24575084"  
/translation="MAVLAATSPGGRLTVFLQVGLLLACAYGLRLGRLGLPL  
VGLTAGVILGPTLLGSLGRLPREDISOHLDDAFQCGVLLVLAIGAPDP  
RILKRGADLAVSLAGLLPLGLGIAHGLVPSALLADSGRGVZALFLGVMCTA  
LPVIAKTLADNLTHRNQQLIAAFDDAVGMLLALVTLASGAGGFVITTTA  
VTVFAACVAGPPIGRSRSDGSRVPSAVTVGVAVVLYGALTAAAGMALGLV  
FVAGTALLHAPVRLAPLRLTVMAVFAVPLVGLVRLMDLTALAEVLLTGLVLL  
VATKFGKAGAAVARSAGMSRYEGLAGAGNSRGMIEVIALVGLRIGVLDVTFTI  
IVLVALITSVATPLRWASSRIVLEEDETERDRLAGMTEPALSGPAPKAREEK  
TPDS"  
complement (11351. .12835)  
/note="ORF11"  
/codon\_start=1  
/transl\_table=1  
/product="alkylhalidase"  
/protein\_id="AA06656.1"  
/db\_xref="GI:24575085"  
/translation="MDVSAOYDVIVGGPGAGSTVTLVRKGRVLOLEKETFPVQ  
IGESLPSVTHGIAHLGVDELKAAFTIKHGTFKGNANPEWTDFAVSRMGA  
TGAYQVERMFDQLLDNARRHGVRENSDVLDDVLAEDGRVGRVRYKSDSEGRHE  
VGSFVVDASNTGLLYKSGAKREYSFPFNALFGYFNGKRLPKNYSDNPTCF  
EHGFWYIPLSPELTSGAVVNRDSALVQDPEKAMESFTAAFLAEKLSAATRT  
EGPYGLRVRKDWMSYKFWAPGMALVQDAACFDPVFSGHVLAITYSGLLAARSIN  
SCLDSVDETSAFEEFARVRYLGRFYFVLDVGFVDMHSDSYFMOARKISNTADSD  
LQSFVELLGVSSDEFAATARQFSEASDEMAGMTHTAGDADGPHGAVPGKVGCR  
LMEQAARLQARALLGDEAGRPVQGGAASTDLGLGVNPGARS"  
13012. .14079  
/note="ORF10"  
/codon\_start=1  
/transl\_table=1  
/product="dNTP-glucose synthase"  
/protein\_id="AA06657.1"  
/db\_xref="GI:24575086"  
/translation="MKALVLSGGSGTRURPISYAMPKQLVPIACKPVLEVVLNIRML  
DIKEVAIVGWAQBIIEAMGDSRFLRLTYIQEPLGIAHCKVLARDFLDEDDFV  
LYIGIMLDGDLISAGHFLHTRPAARIVVRQPDPAFGVIELDGEGRVLRVYKPR  
EPRSLAAVGVYFFADVRAVDATSPRRGELEITDAIOWLLBQGLFVEARVYDYM  
KDTGRVEDVCEKRMGLRALOVSGEVDPESELGVAVVVEGARVTRSRVVGPAVIG  
ACTVVEDSQIGPVASIGRCTVRASRLSDSIVLDDASILAVSGLHSLIGRGARIAPG  
ARGEARHLRVGDHVQIEIAA"  
complement (14212. .14643)  
/note="Caga"  
/codon\_start=1  
/transl\_table=1  
/product="C-1027 apoprotease"  
/protein\_id="AA06658.1"  
/db\_xref="GI:24575087"  
/translation="MSLRHMSRRARFGVAVASTGLAAASQVAFAPAFSVSPAG  
LSDGQSVSVSGAAGETTYIAQCAPVGGQDACNPATFTTDSAGAAFSFVVRK  
SYTGSTPEGTVPVSDCATACNLGAGNSGLDLGHVLTFG"  
complement (14690. .15922)  
/note="ORF8"  
/codon\_start=1  
/transl\_table=1  
/product="amino transferase"  
/protein\_id="AA06659.1"  
/db\_xref="GI:24575088"  
/translation="MSGGVRLLDLSPTPFSEPLHVRPNVSGRDLRIMERIDGALERL  
WFTNDGPLEVFEARVAELTOVRCHVANSNATGIVAAKALGIGGDEIVPFTW  
ATAALDWTGAVPVCLEDEGTADVAVHRLIGRTRAILDHDVHVGFRPARIDELTK  
LAEHGLHLIPAAHAPGCTYRSKPIGEGTAELISFOATKFNVSFGGAVTDDAL  
ADRLARVRQGLNAHEITGSGTVAVMHEISAMGLTSDLESADHTTAINRRNRYLQE

CDS

Query Match 19.4%; Score 97; DB 1; Length 85163;  
Best Local Similarity 55.1%; Pred. No. 0.00029;  
Matches 190; Conservative 0; Mismatches 155; Indels 0; Gaps 0;  
CDS  
/note="ORF7"  
VLDLPGVVRPQDPNELSNCOYVIEVDVAVRAGLHRLDELOAVLQRHNVLARVFPSPG  
CHSCPYSRDLARHAPDLPKVALTERVLSLTGTAVGPEVGVCRILLRAVDGSA  
VPEIHSETEGADGRRPAR"  
complement (15919. .16653)  
/note="ORF7"  
154 GGGAGGTACAGGATCACTGTCCGCGAGCGCGGCGAACTCTGTCTCTCCGCAATGTC 213  
Db GGGATGGGAACGATGAATTCCTCCCGCTCCAGAAAGCGCGCTCTTCCGAGCGACC 16758  
214 TGTTCGGCGTGTTCAGCGAGAGAGCGGTAGTCCACCGCTCGCGGCTGAACGCG 273  
Db TGTTCGGTGTAGTTCAGCGAGAGAGGTAGTGTTCGGCTCGTGGCAGCGACCTCC 16818  
274 TCCGGGCTCGCACCGCGGATGTGTCGCGGGGGTGAGCGCGCTCTGTTCGCCCGCTC 333  
Db TCCGAGGAGGACCGCGGATGCGGTTCCCGCGCAGCAGTTTCCCGCTGTGAGGTGGTG 16878  
334 GTGTCCGACACCCAGGAGACAGGTCCGCGACGATCCGCGAGAGTTCGTACCGTGGCG 393  
Db GTGTCCCGCAGACGAGTGTCTGTATCCGTACACCGCGAGCCATCAGCAACTGGTTC 16938  
394 CTCTTCGCGCTCGCGCTTACCCACCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 453  
Db CCCTTGGACGGTGCTCCGTAGCGCGCACCGGTGCGCTCGCGCGCAGACCGCGAAG 16998  
454 AGGCGCAGCAGGTTCGGTGGGATCCCTCGAGTCGCGGCGCGAAC 498  
Db AGCGTACGAGTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 17043  
16999 AGCGTACGAGTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC

RESULT 29

LOCUS AF170880 25617 bp DNA linear BCT 30-MAY-2003  
DEFINITION Streptomyces spheroides novobiocin biosynthetic gene cluster,  
complete sequence.  
ACCESSION AF170880  
VERSION AF170880.3 GI:31212900  
KEYWORDS  
SOURCE Streptomyces caeruleus  
ORGANISM Streptomyces caeruleus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 25617)  
AUTHORS Steffensky,M., Muehlenweg,A., Wang,Z.X., Li,S.M. and Heide,L.  
TITLE Identification of the novobiocin biosynthetic gene cluster of  
Streptomyces spheroides NCIB 11891  
J. Biol. Chem. 275 (28), 21754-21760 (2000)  
JOURNAL Antimicrob. Agents Chemother. 44 (5), 1214-1222 (2000)  
MEDLINE 20233587  
PUBMED 10770754  
REFERENCE 2 (bases 12457 to 14040)  
AUTHORS Steffensky,M., Li,S.M. and Heide,L.  
TITLE Cloning, overexpression, and purification of novobiocin acid  
synthetase from Streptomyces spheroides NCIB 11891  
J. Biol. Chem. 275 (28), 21754-21760 (2000)  
JOURNAL 20347301  
MEDLINE 10801869  
PUBMED  
REFERENCE 3 (bases 1 to 25617)  
AUTHORS Steffensky,M., Muehlenweg,A., Wang,Z., Li,S.M. and Heide,L.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUL-1999) Pharmazeutische Biologie, Pharmazeutisches  
Institut, Universitaet Tuebingen, Auf der Morgenstelle 8, Tuebingen  
72076, Germany  
REFERENCE 4 (bases 1 to 25617)  
AUTHORS Steffensky,M., Muehlenweg,A., Wang,Z., Li,S.M. and Heide,L.  
TITLE Direct Submission  
JOURNAL Submitted (20-DEC-2000) Pharmazeutische Biologie, Pharmazeutisches  
Institut, Universitaet Tuebingen, Auf der Morgenstelle 8, Tuebingen  
72076, Germany

CDS

CDS

CDS



```

REMARK      Sequence update by submitter
REFERENCE    5 (bases 1 to 25617)
AUTHORS      Steffensky M., Muehlenweg, A., Wang, Z., Li, S.M. and Heide, L.
TITLE        Direct Submission
JOURNAL      Submitted (30-MAY-2003) Pharmazeutische Biologie, Pharmazeutisches
              Institut, Universitaet Tuebingen, Auf der Morgenstelle 8, Tuebingen
              72076, Germany
REMARK      Sequence update by submitter
COMMENT      On May 30, 2003 this sequence version replaced gi:11921116.
FEATURES     Location/Qualifiers
             1..25617
             /organism="Streptomyces caeruleus"
             /mol_type="genomic DNA"
             /strain="NCIB 11891"
             /db_xref="taxon:195949"
             1..1908
             /gene="novA"
             /gene="novA"
             1..1908
             /notes="ABC transporter"
             /codon_start=1
             /transl_table=11
             /product="NovA"
             /protein_id="AAF67494.1"
             /db_xref="GI:7688707"
             /translation="MKSALSTWKPSDRPDPPLPBPQAQLRIIFRLFRPYRGLAVWG
             LIVGASLVAVASPFMRILPTAIPQRTGLSLALGMILTAVLSSVFTVATSLIS
             TTVGQSVMLDRTAVQAQRMPLAFPTTRTGEVQSRIANDIGMQATVTSATSLV
             SNLTAVTAMALMDLWLTVVSLLLPVFVWIGRRVGRKRKITLOROKOMATMAA
             TWTESLVSIGILLGTMGADSLTSPFAESRLVDLEVRSMAGRWMSVIGIVWAA
             MRAVITWAGFVLSGGTVSIGTLVAFVSLQGLFRPAVSLITPGVQMQLSLAFQR
             IFEYLDLPITPERPVALDKVRGEVDFDVFSTEEKDGNLHGLDLTPAGGSLA
             VVGTSQSTLSYLVRPLYDVTVGGVLLDGVDRDLAFDLTARAVGVVVSQETYLPHA
             SVADNIRFAKPATDEIEKAARAQIHEHIVLPDGYDTLVGERGYPSGQKORLA
             IARTILRDPVLVDATSLDTRTEHVAQQAIIDSLSEGRITITIAHRLSTVRDADQI
             VVLDAGRIARGTHIELIDRCRYAALVRDGPAPAPVPFARDGKRRRCVCGVRRAS
             RRSGRDRRAGS"
             2160..3014
             /gene="novB"
             2160..3014
             /gene="novB"
             /notes="similar to aminodesoxychorismate lyase"
             /codon_start=1
             /transl_table=11
             /product="NovB"
             /protein_id="AAF67495.1"
             /db_xref="GI:7688708"
             /translation="MNEPRRAASGRGPLTRRGRVYLIVAAVAAVAVLVLPLRT
             GGTPEGPKKTVSLVPEGRRAQVYDVKVLGVPPTAKTATGARLEPCAAHNN
             PEGYLPATVPVDEGATPAGLLGYMDTARQFSGRITEGARRNGSVYQVTFIASI
             VQAEADTPADMGRVARVNNRLADMDALQMDSTLVNALARSTLDTHTSPKTASPYNT
             YESKGLPPTPIGNPGEAMNAINPTRGDNLIFVTVPAGDTFTADYAEHRNVRFEFN
             ANRSTPTD"
             complement(3030..4088)
             /gene="novC"
             complement(3030..4088)
             /gene="novC"
             /notes="putative oxidoreductase"
             /codon_start=1
             /transl_table=11
             /product="NovC"
             /protein_id="AAF67496.1"
             /db_xref="GI:7688709"
             /translation="MDVWVITGAGQGLSSAYHLRRSLVDPDGTSSPSTTPRPGGAM
             QFRWDSITVEKIHGMALPGMELTGADGSPSSSEVIGGYTAEKEFDLRVHRPVDTV
             AVREGGGLRVETSGVWSARALINATGTWDRPFPYRPGQETFRGRLHTADYFGP
             GEFAGLRVVVGAGSGTQHLMEIADVAATITWTRRPVFRREGPGGEEGGAVALV
             DERVARLPPRSVVSTGLVNEAITRARERGVLDRLPMFDRITPSGVVWDDGRHVDA
             DVLWATGFPAPADHPLAPLRCASPAAGSGWRTRVPYATSATSWNSGNSASTIGANRAG
             RAAVREIRALLERDPLV"
             4196..4627
             /gene="novD"
             4196..4627
             /gene="novD"
             /notes="similar to LmbU protein from Streptomyces
             lincolnensis"
             /codon_start=1
             /transl_table=11
             /product="NovD"
             /protein_id="AAF67497.1"
             /db_xref="GI:7688710"
             /translation="MDLMSDATFTTIVSLTGTDETVTLTRVCEQFLHESAAGRDGL
             LNFVPHATAGVAVIETGAGSDDDLLAALRTLLPADNRWRHRGSPGHGRDHVLPAV
             PPHATLPVIAGRLIELGTWQSVCLVDNTDNPDRYVRLSFLG"
             4808..5461
             /gene="novE"
             4808..5461
             /gene="novE"
             /notes="similar to LmbU protein from Streptomyces
             lincolnensis"
             /codon_start=1
             /transl_table=11
             /product="NovE"
             /protein_id="AAF67498.1"
             /db_xref="GI:7688711"
             /translation="MVASGRASKRGNGATPVRPTAGDATPVDSGQPSDTTYGGLV
             SASRETLRIPRDLISLEAWCRLGGRLLAVCDSSVWVIGDMLVFGONXYGDRYRAMKET
             KLDYQTRNVAWVARKEPSPRRRDSLTFOHMEVAALSSEAQDHLWDPAVLRANKRNE
             LRKQIRASLSGEEDDLRCVQLNLQDLDELRLERREARAGSNLTLDWILSVVDGAV"
             5536..6624
             /gene="novF"
             5536..6624
             /gene="novF"
             /notes="putative oxidoreductase; similar to prephenate
             dehydrogenase"
             /codon_start=1
             /transl_table=11
             /product="NovF"
             /protein_id="AAF67499.1"
             /db_xref="GI:7688712"
             /translation="WRTAVIIGTWICTIGLARLKQGVDSYLMDSPTVALRIAEEVG
             AGTAEPPEPVDLAVVAPVPPVAPVIAASHQSRGTARFYVDVAGVKYSTRELDALQC
             DLATVGGHELTVGRPGSPGLAARGDLFDGRFPAWLPVAVTDAALNRLALVELVACGAI
             PVLDAEADHRAIALGTLPQIALTLVAARLTADSGALRLGLGSVWSEIIPQLVGVDSA
             TSWTQVLAANAAPVVGLEKLSLDLSLETLRGADGDSLAEPDGRLLLEFIQRGID
             GSNRPGRYGIPTETALADVDSVDDRPABLARLFDVAGAGVVMRGIDISQRPDSD
             RTVTISVTPRDAENLLHELRRKWPANS"
             6730..7686
             /gene="novG"
             6730..7686
             /gene="novG"
             /notes="putative regulatory protein"
             /codon_start=1
             /transl_table=11
             /product="NovG"
             /protein_id="AAF67500.1"
             /db_xref="GI:7688713"
             /translation="MTNSGDEEITPASLKATRKGRVSIQSLPPSELVRSGETEHI
             RVLAETDELPPIVVRGTRVVDGMRHWAARFGDESIEVVFVDSPADVFFVLAWE
             LNRAGLPLTLDERKSAQAQIMDSHPHWSRDKIARTTAASTVASLRSSSTACTVGR
             RTGQGRSRNDGTGDRQAAALNAPNASLREVTRAAGISVGTASDVRARLRGEP
             LTAQAQAVMKLREAAQRSGPDYGVLENLRKDPISLRTDLGRLLRLLDGSPGVS
             EQIAQIAGVPEHCRTVTVDMARECAAWQHLADLRDTA"
             7882..9684
             /gene="novH"
             7882..9684
             /gene="novH"
             /notes="similar to peptide synthetase"
             /codon_start=1
             /transl_table=11
             /product="NovH"
             /protein_id="AAF67501.1"
             /db_xref="GI:7688714"
             /translation="WFNTRANKASDQPTPTTESATLAEIMWTRVRRSPSPAIWTNG
             ETLSDVEWNAARNLRLILLDEGAGPGRVALALPRSHLWISVLAVAKAGAVFLPDL
             VNHPRRLSYQLADARPAALLCTVRSAAARLPDGIEMPRVLLDSPERTAVLDALPDIDL

```

```

TDBRGGLAATLAVIYTSSTGRPKGVALTGAGLPALAAKVAAMRVTDGSRVLQ
FASPGDRLVTELLAAFTAGTAVTDTLAGDPLRLALDRGVSHVLRPRRSATM
SPDAPDLRVLVVAGACPAGLVWRWAPGRLLINAYPTCTVCAATMTGPLEPTDEVT
IGRPIGVSVYIILADSRPAAGCEIGELYLSGAGLAQGYLNSPDLTAQMFVNPFAAD
GERMYRTGDLASRDADILPHGRIIDQVELRGRFVRLGEVSVLSQHPDVAQAAAL
WTPAEGPOLATVVVPAGTTPSAGELRHAGRFELDFWVPSAFTTIDAVLTPGKT
DLRAGDPKATPAGGAGTTPAEKVLCDIFRDLDELVDVRSNFFENGNSILLVD
LLOAQEAGLTLMPRTVIDHPTIEQLAAIATLEE"
/gene="novi"
/locus="novi"
/notes="cytochrome P-450"
/codon_start=1
/transl_table=11
/product="Novi"
/protein_id="AA067502.1"
/db_xref="GI:7688715"

Query Match 19.0%; Score 95; DB 1; Length 25617;
Best Local Similarity 54.0%; Pred. No. 0.00073; Mismatches 165; Indels 0; Gaps 0;
Matches 194; Conservative 0;

QY 140 TCAGTGTACCGGGGACGTACAGGATCCACTTCCGCCAGCCCGGGGAACTCCTGCT 199
Db 24019 TCACCCCGCGTTCGGGTATGGGAACGATGAATCGCGCGCGCTTCTCCAGGAATTCCTTCT 23960

QY 200 CTTTCGCCATGATCTCGTCCGCGTGGTTCAGGCGAAGAGCAGCGCTAGTCCACCGCT 259
Db 23959 GCTTGTGATGATCTCCGGTCCGTAGTTCACAGGCAACAGCAGGTAGTGTCCCGTACAT 23900

QY 260 CGGGCGTGAACCGCTCCGGGTGGCCACCGGGATGTGGTCCGGGGTGGAGCGGCGCT 319
Db 23899 GACCGCGCGCTGTTCGGGTGACCAAGATTCGGAAATTCGGACCCCGGAAGTATTTCCCT 23840

QY 320 GCTTCGCCCGGTCTGTGTCGACACACCCAGGAGACAGGTCCGGACCGATGCCGAGAAGT 379
Db 23839 GCTTCAGAAACGTTGTGTCGTCGAGAACCGGATCTCTGTATGTCCGAGCCCGCAGCG 23780

QY 380 TCGTCAAGTGGCGCTCTTCGCGTCCGCGTGGTTCGCGTCCGACCAACCGCTTGGCTCGGCT 439
Db 23779 CCAGCAGTGGCTGCGCTTGGTTCGGTCCGCGTATCCCGCCACCGTCTTGGCGTCCGCCA 23720

QY 440 TCAGCGAGTTCAGCAGGCGAGCAGGTCCGTGGGATGCCCTCCAGCTCCGCGCGGCAAC 498
Db 23719 CCACCGCGGTACAGTTCGGGTCTGTGGCGCACCGCTTCCGTACGATCGCGCAAC 23661

RESULT 31
AX707115/c 31248 bp DNA linear PAT 04-APR-2003
LOCUS
DEFINITION Sequence 1 from Patent WO03014352.
ACCESSION AX707115
VERSION AX707115.1 GI:29563420
KEYWORDS Streptomyces rishiriensis
SOURCE Streptomyces rishiriensis
ORGANISM Streptomyces rishiriensis; Actinobacteridae; Actinomycetales;
REFERENCE 1
AUTHORS Heide, L. and Li, S.M.
TITLE Nucleic acids for aminocoumarin biosynthesis
JOURNAL Patent: WO 03014352-A 1 20-FEB-2003;
Universitaet Tuebingen (DE)
FEATURES
source 1. 31248
/organism="Streptomyces rishiriensis"
/mol_type="unassigned DNA"
/db_xref="taxon:68264"

ORIGIN
Query Match 18.8%; Score 94; DB 6; Length 31248;
Best Local Similarity 53.9%; Pred. No. 0.00093;
Matches 193; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 140 TCAGTGTACCGGGGACGTACAGGATCCACTTCCGCCAGCCCGGGGAACTCCTGCT 199
Db 29686 TCACCGCGGTTCGGGTATGGGAACATGAACCGCGCGCGCTTCCAGGAAGATTTCT 29627

QY 200 CTTTCGCCATGATCTCGTCCGCGTGGTTCAGGCGAAGAGCAGCGCTAGTCCACCGCT 259
Db 29626 CTTTGTGATGATCTCCGACGCGTGTTCAGGCGCAACAGCAGGTAGTGTCCGTACGT 29567

QY 260 CGGGCGTGAACCGCTCCGGGTGGCCACCGGGATGTGGTCCGGGGTGGAGCGGCGCT 319
Db 29566 GGTTCGCTTGTCTTCGGCGCACAGATCGGAATCTGGAGCCCGGTAGGAGTTTCCCT 29507

QY 320 GCTTCGCCCGGTCTGTGTCGACACACCCAGGAGACAGGTCCGACCGATGCCCGAGAAGT 379
Db 29506 GCTTCAGGGTGTGTGTCTGCTCGAGAACCGGATCTCTGTATGTCCCGCCGCGATCGCG 29447

QY 380 TCGTCAAGTGGCGCTCTTCGCGTCCGCGTGGTTCGCGTCCGACCAACCGCTTGGCTCGGCT 439
Db 29446 TCAGCAGTGGCGCTTCTGGTTCGGTCCGCGCTATCCAGCACTGTCTTCCGCTTGCGCA 29387
```

```

TDBRGGLAATLAVIYTSSTGRPKGVALTGAGLPALAAKVAAMRVTDGSRVLQ
FASPGDRLVTELLAAFTAGTAVTDTLAGDPLRLALDRGVSHVLRPRRSATM
SPDAPDLRVLVVAGACPAGLVWRWAPGRLLINAYPTCTVCAATMTGPLEPTDEVT
IGRPIGVSVYIILADSRPAAGCEIGELYLSGAGLAQGYLNSPDLTAQMFVNPFAAD
GERMYRTGDLASRDADILPHGRIIDQVELRGRFVRLGEVSVLSQHPDVAQAAAL
WTPAEGPOLATVVVPAGTTPSAGELRHAGRFELDFWVPSAFTTIDAVLTPGKT
DLRAGDPKATPAGGAGTTPAEKVLCDIFRDLDELVDVRSNFFENGNSILLVD
LLOAQEAGLTLMPRTVIDHPTIEQLAAIATLEE"
/gene="novi"
/locus="novi"
/notes="cytochrome P-450"
/codon_start=1
/transl_table=11
/product="Novi"
/protein_id="AA067502.1"
/db_xref="GI:7688715"

Query Match 19.0%; Score 95; DB 1; Length 25617;
Best Local Similarity 54.0%; Pred. No. 0.00073; Mismatches 165; Indels 0; Gaps 0;
Matches 194; Conservative 0;

QY 140 TCAGTGTACCGGGGACGTACAGGATCCACTTCCGCCAGCCCGGGGAACTCCTGCT 199
Db 24019 TCACCCCGCGTTCGGGTATGGGAACGATGAATCGCGCGCGCTTCTCCAGGAATTCCTTCT 23960

QY 200 CTTTCGCCATGATCTCGTCCGCGTGGTTCAGGCGAAGAGCAGCGCTAGTCCACCGCT 259
Db 23959 GCTTGTGATGATCTCCGGTCCGTAGTTCACAGGCAACAGCAGGTAGTGTCCCGTACAT 23900

QY 260 CGGGCGTGAACCGCTCCGGGTGGCCACCGGGATGTGGTCCGCGTCCGACCAACCGCTTGGCTCGGCT 319
Db 23899 GACCGCGCGCTGTTCGGGTGACCAAGATTCGGAAATTCGGACCCCGGAAGTATTTCCCT 23840

QY 320 GCTTCGCCCGGTCTGTGTCGACACACCCAGGAGACAGGTCCGGACCGATGCCGAGAAGT 379
Db 23839 GCTTCAGAAACGTTGTGTCGTCGAGAACCGGATCTCTGTATGTCCGAGCCCGCAGCG 23780

QY 380 TCGTCAAGTGGCGCTCTTCGCGTCCGCGTGGTTCGCGTCCGACCAACCGCTTGGCTCGGCT 439
Db 23779 CCAGCAGTGGCTGCGCTTGGTTCGGTCCGCGTATCCCGCCACCGTCTTGGCGTCCGCCA 23720

QY 440 TCAGCGAGTTCAGCAGGCGAGCAGGTCCGTGGGATGCCCTCCAGCTCCGCGCGGCAAC 498
Db 23719 CCACCGCGGTACAGTTCGGGTCTGTGGCGCACCGCTTCCGTACGATCGCGCAAC 23661

RESULT 30
AX707121/c 25617 bp DNA linear PAT 04-APR-2003
LOCUS
DEFINITION Sequence 7 from Patent WO03014352.
ACCESSION AX707121
VERSION AX707121.1 GI:29563426
KEYWORDS Streptomyces caeruleus
SOURCE Streptomyces caeruleus
ORGANISM Streptomyces caeruleus; Actinobacteridae; Actinomycetales;
REFERENCE 1
AUTHORS Heide, L. and Li, S.M.
TITLE Nucleic acids for aminocoumarin biosynthesis
JOURNAL Patent: WO 03014352-A 7 20-FEB-2003;
Universitaet Tuebingen (DE)
FEATURES
source 1. 25617
/organism="Streptomyces caeruleus"
/mol_type="unassigned DNA"
/db_xref="taxon:195949"

ORIGIN
Query Match 19.0%; Score 95; DB 6; Length 25617;
Best Local Similarity 54.0%; Pred. No. 0.00073;
Matches 193; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 140 TCAGTGTACCGGGGACGTACAGGATCCACTTCCGCCAGCCCGGGGAACTCCTGCT 199
Db 24019 TCACCCCGCGTTCGGGTATGGGAACGATGAATCGCGCGCGCTTCTCCAGGAATTCCTTCT 23960

QY 200 CTTTCGCCATGATCTCGTCCGCGTGGTTCAGGCGAAGAGCAGCGCTAGTCCACCGCT 259
Db 23959 GCTTGTGATGATCTCCGGTCCGTAGTTCACAGGCAACAGCAGGTAGTGTCCCGTACAT 23900

QY 260 CGGGCGTGAACCGCTCCGGGTGGCCACCGGGATGTGGTCCGCGTCCGACCAACCGCTTGGCTCGGCT 319
Db 23899 GACCGCGCGCTGTTCGGGTGACCAAGATTCGGAAATTCGGACCCCGGAAGTATTTCCCT 23840

QY 320 GCTTCGCCCGGTCTGTGTCGACACACCCAGGAGACAGGTCCGGACCGATGCCGAGAAGT 379
Db 23839 GCTTCAGAAACGTTGTGTCGTCGAGAACCGGATCTCTGTATGTCCGAGCCCGCAGCG 23780

QY 380 TCGTCAAGTGGCGCTCTTCGCGTCCGCGTGGTTCGCGTCCGACCAACCGCTTGGCTCGGCT 439
Db 23779 CCAGCAGTGGCTGCGCTTGGTTCGGTCCGCGTATCCCGCCACCGTCTTGGCGTCCGCCA 23720

QY 440 TCAGCGAGTTCAGCAGGCGAGCAGGTCCGTGGGATGCCCTCCAGCTCCGCGCGGCAAC 498
Db 23719 CCACCGCGGTACAGTTCGGGTCTGTGGCGCACCGCTTCCGTACGATCGCGCAAC 23661
```

440 TGAGCGAGTTTCAGCAGGCGGACACAGCTCGTGCAGATGCCCTTCGACGCTCGGCGCGGAA 497

Db 29386 CCAAGCTGCGACCAAGTTCCTCCGGATCTCTGCGCGACCCCGTTCGTAAGCTCGGCGGAA 293329

RESULT 32

AF235050/c

LOCUS AF235050.3

DEFINITION Streptomyces rishiriensis strain DSM 40489 coumermycin A1 biosynthetic gene cluster, complete sequence.

ACCESSION AF235050

VERSION AF235050

KEYWORDS GI:38147033

SOURCE Streptomyces rishiriensis

ORGANISM Streptomyces rishiriensis

Bacteria; Actinobacterii; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 35359)

AUTHORS Wang, Z.-X., Li, S.-M. and Heide, L.

TITLE Identification of the coumermycin A(1) biosynthetic gene cluster of Streptomyces rishiriensis DSM 40489

MEDLINE Antimicrob. Agents Chemother. 44 (11), 3040-3048 (2000)

PUBMED 20493109

REFERENCE 2 (bases 1 to 35359)

AUTHORS Wang, Z.-X., Li, S.-M. and Heide, L.

TITLE Direct Submission

JOURNAL Submitted (16-FEB-2000) Pharmazeutische biologie, Pharmazeutisches Institut, University of Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany

REFERENCE 3 (bases 1 to 35359)

AUTHORS Wang, Z.-X., Li, S.-M. and Heide, L.

TITLE Direct Submission

JOURNAL Submitted (30-MAR-2000) Pharmazeutische biologie, Pharmazeutisches Institut, University of Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany

REMARK Amino acid sequence updated by submitter

REFERENCE 4 (bases 1 to 35359)

AUTHORS Wang, Z.-X., Li, S.-M. and Heide, L.

TITLE Direct Submission

JOURNAL Submitted (31-OCT-2000) Pharmazeutische biologie, Pharmazeutisches Institut, University of Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany

REMARK Sequence update by submitter

REFERENCE 5 (bases 1 to 35359)

AUTHORS Wang, Z.-X., Li, S.-M. and Heide, L.

TITLE Direct Submission

JOURNAL Submitted (06-JUL-2001) Pharmazeutische biologie, Pharmazeutisches Institut, University of Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany

REMARK Sequence update by submitter

REFERENCE 6 (bases 1 to 35359)

AUTHORS Wang, Z.-X., Li, S.-M. and Heide, L.

TITLE Direct Submission

JOURNAL Submitted (06-JUN-2002) Pharmazeutische biologie, Pharmazeutisches Institut, University of Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany

REMARK Sequence update by submitter

COMMENT On Nov 3, 2003 this sequence version replaced gi:14625486.

FEATURES

source 1..35359

/organism="Streptomyces rishiriensis"

/mol\_type="genomic DNA"

/strain="DSM 40489"

/db\_xref="taxon:68264"

<1..1028

/note="ORF3"

/codon\_start=3

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AA011882.1"

/db\_xref="GI:38147034"

CDS

5929. 7728  
/gene="coulH"  
/note="coulH"  
5929. 7728  
/gene="coulH"  
/note="coulH"  
/codon\_start=1  
/transl\_table=11  
/product="acyl-CoA synthetase"  
/protein\_id="AAG29780.1"  
/db\_xref="GI:11095214"  
/translation="MSEKGLNKASDRSATRTASATLAEIWEPTVSRSSPAIVNS  
ETUSYDEVNARQNLAKLLSEGPGRVVALALPRSSRWISVLAVAKAGAAFLPVD  
VNYPEQRIAYMLDAEPALLCTIRTAVKPLPADIGVPLQVLDSAEQATLADLPDIDV  
TEBERGSLATLAIATVYITSGSTGPKGVALTGAGLPALAAKAVTMQVTEBSVLQ  
FASPSFAYTELIAATVPGSDTLVGDPLAKVLRQGVSHAVLPAAVATM  
SPDAAPDLRVLAGECAGLVEQMSPGRLLVNAYGPTCTCAVATGTLPTDVT  
IGRPIGVSHIILDELPAVAGEIGELYSGAGLARGYLNRPDVTAHVFNAPRAVL  
GERMYRGDLASRADZILFHRIGIDQVELGRFRIELGEVESVLQSDHQAQAVL  
RAGDDQQLAYVYPHTGTPAGELREHAGRYLPIDYWPFSVYATIDALIPVTPSKT  
DRAALPIKTRTAPAGQKPTSAEKILFDIFRDLFGLEIDVRSNFFEMGNSILAVDL  
IQRAQAGLVLLPRTVIDHPTIEQAAIANLGE"  
7732. 8955  
/gene="coulH"  
/note="coulH"  
7732. 8955  
/gene="coulH"  
/note="coulH"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P-450"  
/protein\_id="AAG29781.1"  
/db\_xref="GI:11095215"  
/translation="MSTRPMVSADELDLSPILHAELYELDAVFRHLRANRPMYQ  
QPRGEQFVWISYADNVEYKDKAHTFTENGALATLITGGDSAGLAVITDGVYR  
HQVNLISLKSFGPOMELIATNSLRVETDGLLALDRGEDCAARDISAKVPLGACD  
LLEIPQDREYLLGTAHWSTYADPEESWAKNEIILYFKLKERRRGDRDM  
VSLLANGLDHPHAAEQIANCYGLMIGDETGRAITGTALIANPQWGLKNG  
DVLKATTEARLWTPSLGGRKATGVVINGQIKAGEWVSWISSANRDETVD  
ADEKPLATPNKHTFAYGSHYCLHGLRMEVAVLDGLRRLVGLDLEQIEERWLYS  
SILHGMSPILKITA"  
9115. 9891  
/gene="coulH"  
/note="coulH"  
9115. 9891  
/gene="coulH"  
/note="coulH"  
/note="similar to 3-ketoacyl-[ACP]-reductase"  
/codon\_start=1  
/transl\_table=11  
/product="CoulH"  
/protein\_id="AAG29782.1"  
/db\_xref="GI:11095216"  
/translation="MTGPAASAVACEPVMVVTGAGRGIGATATRLAAGWAVIVDR  
VEQDTATVSAIRAAGRASGLGDSVAQAVTATVASAVEEFGHLDVLVNCAGITOD  
RLLLTMSDOENRDLVNLGSMRCSFVAGHMERQGHRIINPSSVAARGNAGOTNY  
ATAKAIAGFTRTLAEPGPGVTVNSIAPGFVATPMVAELTERLGRVDRDTVLLEAM  
SSAVGRIGTPBEILAAVAVFTRPESGYLTGETIHVEGRR"  
9888. 10625  
/gene="coulH"  
/note="coulH"  
9888. 10625  
/gene="coulH"  
/note="coulH"  
Query Match 18.8%; Score 94; DB 1; Length 35359;  
Best Local Similarity 53.9%; Pred. No. 0.0009;  
Matches 193; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 140 TCACGTGTACGGGGGACGTACAGGATCCACTGTCCGCCAGCCCGGGGAACTCCTGCT 199  
DB 33797 TCACGGCGGTTTCGGGTATGGGAACAATGAAACGGCGCCCTTCCAGGAACGACTTCT 33738  
QY 200 CTTTCCCATGATCTCGTCGGGTGGTTCAGGCGAAGAGCAGCGGTAGTCCACCGGCT 259  
DB 33737 CTTTGGTGATGATCTCCGACGCGTAGTTCAGGCCAACAGCAGGTAGTCCGGTACGT 33678  
QY 260 CGSGCGTGAACCGGTCCGGGGTGGCCACCGGATGTGCGTCCGGGGGTAGCCCGCCCT 319  
DB 33677 GGTTCGCTTCTGTTTCGGCGACCATGCGAATCTCGAGCCCGTAGGAGTTTTCCT 33618  
QY 320 GCTTGGCCGGCGTGTGTGTCGCACACCCAGGAGACAGGTCCCGACCGATGCGCAGAGT 379  
DB 33617 GCTTCAGGGCTGTGTGTGTCGCTGCAGAACCGGATCTCTGATGTCCACGCCCGATGCGG 33558  
QY 380 TCGTCACGTTGGCGCTTTCGGCGCTGCGCGCGGTAGCCACACCGCTTGCCTCGGCT 439  
DB 33557 TCAGCAGTGGCTGCTTGTGTCGGTGGCGGTATCCAGCCACTGTCTTTCGGTCTGCCA 33498  
QY 440 TGACGAGTTTACAGAGGGGAGCAGGTGCGTCCGATGCCCTCGACGTCGGCGGGAA 497  
DB 33497 CCAAGCTCGGCACCACTTCCCGATCTGTGCGCGCACCGCTTCCGTACGCTCGGCGAA 33440  
RESULT 34  
AX574040/c

DB 33737 CTTTGGTGATGATCTCCGACCGTAGTTCAGGCGCAACAGCAGGTAGTCCGGTACGT 33678  
QY 260 CGSGCGTGAACCGGTCCGGGGTGGCCACCGGATGTGCGTCCGGGGGTAGCCCGCCCT 319  
DB 33677 GGTTCGCTTCTGTTTCGGCGACCATGCGAATCTCGAGCCCGTAGGAGTTTTCCT 33618  
QY 320 GCTTGGCCGGCGTGTGTGTCGCACACCCAGGAGACAGGTCCCGACCGATGCGCAGAGT 379  
DB 33617 GCTTCAGGGCTGTGTGTGTCGCTGCAGAACCGGATCTCTGATGTCCACGCCCGATGCGG 33558  
QY 380 TCGTCACGTTGGCGCTTTCGGCGCTGCGCGCGGTAGCCACACCGCTTGCCTCGGCT 439  
DB 33557 TCAGCAGTGGCTGCTTGTGTCGGTGGCGGTATCCAGCCACTGTCTTTCGGTCTGCCA 33498  
QY 440 TGACGAGTTTACAGAGGGGAGCAGGTGCGTCCGATGCCCTCGACGTCGGCGGGAA 497  
DB 33497 CCAAGCTCGGCACCACTTCCCGATCTGTGCGCGCACCGCTTCCGTACGCTCGGCGAA 33440  
RESULT 33  
AX707116/c  
LOCUS AX707116 35359 bp DNA linear PAT 04-APR-2003  
DEFINITION Sequence 2 from Patent WO03014352.  
ACCESSION AX707116  
VERSION AX707116.1 GI:29563421  
KEYWORDS Streptomyces rishiriensis  
SOURCE Streptomyces rishiriensis  
ORGANISM Streptomyces rishiriensis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Heide, L. and Li, S. M.  
TITLE Nucleic acids for aminocoumarin biosynthesis  
JOURNAL Patent: WO 03014352-A 2 20-FEB-2003;  
Universitaet Tuebingen (DE)  
FEATURES  
source  
1. 35359  
/organism="Streptomyces rishiriensis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:68264"  
ORIGIN  
Query Match 18.8%; Score 94; DB 6; Length 35359;  
Best Local Similarity 53.9%; Pred. No. 0.0009;  
Matches 193; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 140 TCACGTGTACGGGGGACGTACAGGATCCACTGTCCGCCAGCCCGGGGAACTCCTGCT 199  
DB 33797 TCACGGCGGTTTCGGGTATGGGAACAATGAAACGGCGCCCTTCCAGGAACGACTTCT 33738  
QY 200 CTTTCCCATGATCTCGTCGGGTGGTTCAGGCGAAGAGCAGCGGTAGTCCACCGGCT 259  
DB 33737 CTTTGGTGATGATCTCCGACGCGTAGTTCAGGCCAACAGCAGGTAGTCCGGTACGT 33678  
QY 260 CGSGCGTGAACCGGTCCGGGGTGGCCACCGGATGTGCGTCCGGGGGTAGCCCGCCCT 319  
DB 33677 GGTTCGCTTCTGTTTCGGCGACCATGCGAATCTCGAGCCCGTAGGAGTTTTCCT 33618  
QY 320 GCTTGGCCGGCGTGTGTGTCGCACACCCAGGAGACAGGTCCCGACCGATGCGCAGAGT 379  
DB 33617 GCTTCAGGGCTGTGTGTGTCGCTGCAGAACCGGATCTCTGATGTCCACGCCCGATGCGG 33558  
QY 380 TCGTCACGTTGGCGCTTTCGGCGCTGCGCGCGGTAGCCACACCGCTTGCCTCGGCT 439  
DB 33557 TCAGCAGTGGCTGCTTGTGTCGGTGGCGGTATCCAGCCACTGTCTTTCGGTCTGCCA 33498  
QY 440 TGACGAGTTTACAGAGGGGAGCAGGTGCGTCCGATGCCCTCGACGTCGGCGGGAA 497  
DB 33497 CCAAGCTCGGCACCACTTCCCGATCTGTGCGCGCACCGCTTCCGTACGCTCGGCGAA 33440  
RESULT 34  
AX574040/c

LOCUS AX574040 1290 bp DNA linear PAT 07-JAN-2003  
DEFINITION Sequence 120 from Patent WO02079505.  
ACCESSION AX574040  
VERSION AX574040.1 GI:27551597  
KEYWORDS  
SOURCE  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.  
REFERENCE  
1 Farnet, C.M., Zazopoulos, E. and Staffa, A.  
Compositions and methods for identifying and distinguishing  
orthosomycin biosynthetic loci  
Patent: WO 02079505-A 120 10-OCT-2002;  
Ecopia Biosciences Inc. (CA)  
FEATURES  
Location/Qualifiers  
1..1290  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"  
ORIGIN  
Query Match 18.5%; Score 92.4; DB 6; Length 1290;  
Best Local Similarity 53.6%; Pred. No. 0.0037; Indels 0; Gaps 0;  
Matches 192; Conservative 0; Mismatches 166;  
142 ACCTGTACGCGGGGACGTACAGATCCACTGTCCGCCAGCCGCGGAACTCTCTGCTCC 201  
Db 1250 ACCCGAGGTGGCGGAGGAGATCAGCTTGTTCGGCCGCCACCGTGGCCGCTCG 1191  
QY 202 TTGCGCATGATCTCTGTCGGGGTGTTCAGGCGAAGACGAGCGGTAGTCCACCGGTG 261  
Db 1190 CGCGCCACCATCTCTGTCGGAAGTCCACGCGACGACCATGAGTGTGTCGGGGCGGG 1131  
QY 262 GCGGTGAACGCGTCCGGGTGCGCACCGGGATGTGTGTCGGGGGTGAGCGGCCCTGC 321  
Db 1130 GCCCGGAGTGCSCCTCGTGTATGATCTCGATGTCGATACCGAGGTTCGCGCGGACC 1071  
QY 322 TTGCGCGGTCTGTGTGACACCCAGGAGACAGTCCGACCGATGCGCGAGATTC 381  
Db 1070 TTGTCGGGTGTGCTCGCGGCTACGGGATCAGCGTGGTTCGATCCCGCAGTACTGC 1011  
QY 382 GTACAGTGGCGTCTTGGCGGTCGCGGTAGCCACACCGCTTGCCTCGGCTTG 441  
Db 1010 AGAGAGGTGTTCCTTGTGGAGGCGCGGTAGACGTGACGGTGTTCGCTGTGCG 951  
QY 442 AGCGAGTTCAGAGGCGAGCAGGTGCGTGGGATGCGCTTCGACGTGCGCGGAAAC 499  
Db 950 AGACCATGCAGCATCTTACGAGTTCGTCGGGTGCGCGCGCGCAGCGTTGTCGCGAAC 893  
RESULT 35  
AX204973/c 14252 bp DNA linear PAT 30-AUG-2001  
LOCUS AX204973 Sequence 8 from Patent WO015180.  
DEFINITION AX204973  
ACCESSION AX204973  
VERSION AX204973.1 GI:15394253  
KEYWORDS  
SOURCE  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.  
REFERENCE  
1 Staffa, A., Zazopoulos, E., Mercure, S. and Nowacki, P.  
Genetic locus for evernimycin biosynthesis  
Patent: WO 015180-A 8 02-AUG-2001;  
Ecopia Biosciences Inc. (CA); Farnet, Chris (CA)  
FEATURES  
Location/Qualifiers  
1..14252  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"  
misc\_feature 459..1280

/note="ORF 5 (positive strandedness)"  
1280..2566  
/note="ORF 6 (positive strandedness)"  
2677..3747  
/note="ORF 7 (positive strandedness)"  
3899..4774  
/note="ORF 8 (positive strandedness)"  
4893..5303  
/note="ORF 9 (positive strandedness)"  
5385..6306  
/note="ORF 10 (negative strandedness)"  
6350..7204  
/note="ORF 11 (negative strandedness)"  
7371..8198  
/note="ORF 12 (negative strandedness)"  
8304..9038  
/note="ORF 13 (negative strandedness)"  
9462..10493  
/note="ORF 14 (positive strandedness)"  
10665..11384  
/note="ORF 15 (negative strandedness)"  
11387..12700  
/note="ORF 16 (negative strandedness)"  
12971..14185  
/note="ORF 17 (negative strandedness)"  
ORIGIN  
Query Match 18.5%; Score 92.4; DB 6; Length 14252;  
Best Local Similarity 53.6%; Pred. No. 0.0019;  
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
142 ACCTGTACGCGGGGACGTACAGATCCACTGTCCGCCAGCCGCGGAACTCTCTGCTCC 201  
Db 2529 ACCCGAGGTGGCGGAGGATCAGCTTGTTCGGCCGCCACCGTGGCCGCTCG 2470  
QY 202 TTGCGCATGATCTCTGCGGGTGTTCAGCGGAGAGCAGCGGTAGTCCACCGCTGC 261  
Db 2469 CGCGCCAGATCTCTGTCGGAAGTCCACGCGAGCACCAGGTAGTGTTCGGCGCGG 2410  
QY 262 GCGGTGAACGCGTCCGGGTGCGCACCGGGATGTGTGTCGGGGGTGAGCGGCCCTGC 321  
Db 2409 GCCCGGAGTGGCGCTCGCTGATGATCTCGATGTCGATGTCGAGGGTCCCGGCGG 2350  
QY 322 TTGCGCGCGTGTGTCGACACCCAGAGACAGGTCCGACCGATGCGCGGCGGAACTTC 381  
Db 2349 TTGTCGGGTGCGCTCGCGGCGTACGGATCAGCGTGGGTGATCCCGCAGTACTGC 2290  
QY 382 GTACAGGTGGCGCTTTCGCGGTGCGCGGTAGCCACACCGCTTGGCCCTCGGCTTG 441  
Db 2289 AGAGGGGTGTTCCTTGTGAGGCGCGGTAGACGTGTCACGCGGTGCTTCGCTGCG 2230  
QY 442 AGCGAGTTCAGAGGCGGCGAGCGGTGCGTGGGATGCGCTTCGACGTGCGGCGGAAAC 499  
Db 2229 AGACCATGCAGCATCTTACAGATGTCGCGGTGCGCGCGCGCAGCTTGTTCGCGAAC 2172  
RESULT 36  
AX574200/c 48221 bp DNA linear PAT 07-JAN-2003  
LOCUS AX574200 Sequence 280 from Patent WO02079505.  
DEFINITION AX574200  
ACCESSION AX574200  
VERSION AX574200.1 GI:27551679  
KEYWORDS  
SOURCE  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.  
REFERENCE  
1 Farnet, C.M., Zazopoulos, E. and Staffa, A.  
Compositions and methods for identifying and distinguishing  
orthosomycin biosynthetic loci  
Patent: WO 02079505-A 280 10-OCT-2002;  
Ecopia Biosciences Inc. (CA)



```

6-aminohexanoate-dimer hydrolase (EC 3.5.1.46) protein"
/protein_id="CAC49490.1"
/db_xref="GI:15140977"
/db_xref="GOA:Q92UN4"
/db_xref="SPTREMBL:Q92UN4"
/translation="MLVITLALLGFHSAQAQDSVTTILTRLETRBELKPLKTVV
ARDGTTIAEGRGRTPESTINIKASKSIYSVALVGIADKGLDGPQKIAPLKED
LPSQADRIINDITIGLNSMQAGLGRSPNRYWVSRNWRPFAALQFPDDPFGRM
LYSTATLHLLSAILIKVGGKPTLALAREWLDPEVGRICAWERDFQGIYLGQWAMS
ARSLIAFGLVNRGRATDRGVDPADWVLSWRPRTASRTFGDGYGYWTRIGDE
DYYAWYGGQMLYTVPSLNTLVVMTSDSGPSARNGYRDALHGLVLAIEISTAKAD"
complement(3296..5236)
/gene="lacZ2 OR Smb20966"
complement(3296..5236)
/gene="lacZ2 OR Smb20966"
/EC_number="3.2.1.23"
/function="Small Molecule Metabolism; Degradation; carbon
compounds"
/notes="Product confidence : putative"
Gene name confidence : putative
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative beta-galactosidase protein"
/protein_id="CAC49491.1"
/db_xref="GI:15140978"
/db_xref="GOA:Q92UN3"
/db_xref="SPTREMBL:Q92UN3"
/translation="MCYYPEQPRGRWAEDARMVELGSLVSWRICEFAWKIEPRSGE
FWHEWLEDAIDVLGKAGLVILGTPTAAPPKLVNRYEPIIPVDATGAVRKFGRHHY
CFSSRYSEARAI TEAMARRYGEHIYHWAMTDNEYGDHDTIYSYSAEAVGAPLWL
AERYSGIDLENAWGTFSWMSRYDSEEDIDLENNLVEBPSPTHGVDFIRFSDQVKSF
NKAQVDIIRAHSGRPVTHNFSQNTDFPHRVGEDIDIASWDVPMGLGRLAAK
DKEHVLKVDQDPAHPNDLVRAGVRGVVMEQOPVNNWAHQSPADGWLRLWTW
LAYAGVDWVSIFRRWQAPPAEQFHAGLLPNSADQGYLEVAWVEMKELPEGEV
RGAQVAIVDIFESWATRVLPQKRSISASVALDQYLEVAWVEMKELPEGEV
YKILADLVIAEAPVERLTRAQKVPGRSGSKTDMHPIPELGGPPGLAKLIDIS
VRYLVAANGALLDKVIGDAGWAGLESPLDGLRVTNRGNLRFAPNFRIPAEVPAG
ASAEFYVGGRTLKPDVVAIWR"
complement(5426..6460)
/gene="Smb20967"
complement(5426..6460)
/gene="Smb20967"
/function="MISCELLANEOUS; Not classified regulator"
/notes="Product confidence : putative"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transcriptional regulator, lacI family
protein"
/protein_id="CAC49492.1"
/db_xref="GI:15140979"
/db_xref="GOA:Q92UN2"
/db_xref="SPTREMBL:Q92UN2"
/translation="MORPIKTIARETGLSIATVSKALKHSQVPERTRAVVLDAER
VGYELNHLGQLRTGTYQVAALIMTAPGKRENEWEGVYQAQLSGISWALESPRVS
LYAVRDFESQETISQIVSKKADGIIISGTRADDRIELMQEAFPFVYIGMSIHNA
PHAYVDADNEQMLRLSVARLIERHRRILALNPKQFAYGVVRLSRYKALEKAGLPY
DPAVALHGLTLPAGRENVVTSALADPTTAVICANEATLGAFGSGFHARGLVHGRDA
VINATDNLNVSQYFAPPTITSYLYPIGCEPALLGKFTLRMEGEPPDALQTLAMPDIE
RSDRLRPER"
complement(6468..8180)
/gene="Smb20968"
complement(6468..8180)
/gene="Smb20968"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical"
Gene name confidence : hypothetical
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"

```

```

/protein_id="CAC49493.1"
/db_xref="GI:15140980"
/db_xref="GOA:Q92UN1"
/db_xref="SPTREMBL:Q92UN1"
/translation="MLFDDIHVPFVRKGRFLTSTWRVGRDGERALYLRRVSGGDER
PSLGRCLVEFLDGEGRPAATQFOLSPEQDVARITGRGIVRFVIGDGERLHIRGNAGV
RLVLESRDYVYRTFTGGSCCLVAAGENKFI PRGRGLBLAVSGKWRDSTDVSTFL
SGSGTTEGVNDFPRALPPFEGFSGFEAAHVAVTAEGRATRSIPAGFQGEBAQLAA
YLLWANTVFAEGLVLTTPALPYMSKNGMINIWSMDNAFSAALGVAAPDESELAFGQALFD
HQDASGLLDYNDREALFAFTKPPVHWASVCARENFAFLTPERRAYLRDAIGRQV
SYMLTGRAGDNVLPSPYFHNDSGMDNASFFABGGFVLPDLFPVFLILACEALANLLE
DDPKAARLRIADLEQALLVLTGTGETFAARLAADPERILFGBESLIOPMPLLLASR
LPAMCRQIVARIVEGAFITDMGPATESPRSPYEDDGYWRGPIWAPTVLLWDGLRR
QEGALARELAQKFCALANAHGVAENFDARSGRGLDRAPATSAVYLLLAQSLRDEK
PVKS"
complement(8184..8996)
/gene="Smb20969"
complement(8184..8996)
/gene="Smb20969"

Query Match      18.28; Score 90.8; DB 1; Length 303100;
Best Local Similarity 53.4%; Pred. No. 0.0013;
Matches 191; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy 140 TCACGTGTACGGGGGACGTACAGGATCCACTGTCCGCCAGCCCGCGAACTCTTGGCT 199
      |||||
Db 158066 TCACCGCACATTGCGGAATGCGACGACGAAGCGCGCCGCCAGGACGGATATAGCGCT 158007

Qy 200 CTTTCGCATGATCTCTTCGCGGTGTTCCAGGCGAGACAGCGGTAGTCCACCGCGT 259
      |||||
Db 158006 GGGCGGCGAGCATCTCTCGGTGAGTTCAGGGCAGGATGACACATAGTCGGGTTTAA 157947

Qy 260 CGGCGGTGAACCGCTCCCGGGGTGCGCACCGGGATGTGCGTCCGGGGGTGAGCGCGCCCT 319
      |||||
Db 157946 CGGCTCGATCTCTCGCGGGATCGTAGATCGAATGTGCTCGCGGGCAGAGCTTGCCT 157887

Qy 320 GTTTCGGCGGGGTGCTGTGACACACCAGACAGCAGGTCCGACCATCCCGCAGAGT 379
      |||||
Db 157886 GCTTCAGATCGTTCGCGTCGACGATGAAGTCGATGTCGACGCGGTGAGCCCGCAGACGT 157827

Qy 380 TCGTCACGTGGCGCTCTTTCGCGCTCGCGCGGTACGCCACACCCGCTTGCCTTCGCGCT 439
      |||||
Db 157826 TCAGGAAGGTGTGCGCTTTTTCGCGCGGCGCATAGCGCGCGCTTGTTCGCTTC 157767

Qy 440 TGAACGAGTTTACAGCGGGCGAGCAGGTTCGGTGGGATGCGCTTCGACGTCCGCGCGAA 497
      |||||
Db 157766 TGGCGTCCGCGAGAACCGCGGGAATCCATCGCAGACCGAAGCGATCGCTTGCAGAA 157709

RESULT 38
AX574197/c
LOCUS      AX574197      45055 bp      DNA      linear      PAT 07-JAN-2003
DEFINITION      Sequence 277 from Patent WO02079505.
ACCESSION      AX574197
VERSION        AX574197.1      GI:27551676
KEYWORDS
SOURCE        Streptomyces mobaraensis
ORGANISM      Streptomyces mobaraensis
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS      1 Farnet, C.M., Zazopoulos, E. and Staffa, A.
TITLE        Compositions and methods for identifying and distinguishing
              orthomycin biosynthetic loci
JOURNAL      Patent: WO 02079505-A 277 10-OCT-2002;
              Ecopia Biosciences Inc. (CA)
              Location/Qualifiers
FEATURES
source      1..45055
              /organism="Streptomyces mobaraensis"
              /mol_type="unassigned DNA"
              /db_xref="taxon:35621"
ORIGIN

```

Query Match	18.1%; Score 90.6; DB 6; Length 45055;
Best Local Similarity	52.5%; Pred. No. 0.0023;
Matches 198; Conservative	0; Mismatches 179; Indels 0; Gaps 0;
QY	122 CACCGCGCGGTTCAGTTCGTTCACACCTTCACGCGCGGACGTACAGGATCCACTGTCCGCCAG 181
Db	9974 CCGCGGACCGCGTTCACACCTTCACGCGCGGACGTACAGGATCCACTGTCCGCCAG 9915
QY	182 CCGCGCGGAATCTCGTTCCTTCGCAATGATCTCGCGTGGTTCAGCGGAGGACCA 241
Db	9914 CCGCGCGGAATCTCGTTCCTTCGCAATGATCTCGCGTGGTTCAGCGGAGGACCA 9855
QY	242 GCGGTAGTCCACCGCGTTCGCGGTGAACCGCGTCCGGGTGCGCACCGCGGATGTCGTCG 301
Db	9854 GGAAACCGGTTCGCGGTTCACGCGCGCGGCTTCCTTCGAGACGATTCGCGGATGTCGTCG 9795
QY	302 CGCGGTGAGCGCGGCGCTTCGTTCGCGCGGCTTCGTTCGCGCACCGCGGATGTCGTCG 361
Db	9794 CGCGGTGAGCGCGGCGCTTCGTTCGCGCGGCTTCGTTCGCGCACCGCGGATGTCGTCG 9735
QY	362 GACCGATGCGCGGAGTTCGTTCAGGTGCGGTCTTCGCGTTCGCGCGGTTCAGCGGACCA 421
Db	9734 CGTCAGCGCGGAGTTCGTTCAGGTGCGGTCTTCGCGTTCGCGCGGTTCAGCGGACCA 421
QY	422 CCGGTTCGCGGTTCGCGGTTCAGCGGTTCAGCGGCGGAGTTCGCGTTCGCGGATGCGCT 481
Db	9674 TCGCGAGCGCGGTTCGCGGTTCGCGGAGTTCGCGGAGTTCAGCGGCGGATTCCT 9615
QY	482 CGACGTGCGCGCGGAC 498
Db	9614 CGTTCGCGCGGCGGAC 9598
RESULT 39	
AX574038/c	
LOCUS	AX574038 1224 bp DNA linear PAT 07-JAN-2003
DEFINITION	Sequence 118 from Patent WO02079505.
ACCESSION	AX574038
VERSION	AX574038.1 GI:27551596
KEYWORDS	Streptomyces mobaraensis
SOURCE	Streptomyces mobaraensis
ORGANISM	Streptomyces mobaraensis
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
AUTHORS	Streptomycineae; Streptomycetaceae; Streptomyces.
TITLE	1
JOURNAL	Parnet, C.M., Zazopoulos, E. and Staffa, A.
FEATURES	Compositions and methods for identifying and distinguishing
source	orthomycinin biosynthetic loci
	Patent: WO 02079505-A 118 10-OCT-2002;
	Ecopia Biosciences Inc. (CA)
	Location/Qualifiers
	1..1224
	/organism="Streptomyces mobaraensis"
	/mol_type="unassigned DNA"
	/db_xref="taxon:35621"
ORIGIN	
Query Match	17.6%; Score 87.8; DB 6; Length 1224;
Best Local Similarity	52.6%; Pred. No. 0.015;
Matches 191; Conservative	0; Mismatches 172; Indels 0; Gaps 0;
QY	136 CTGTCACGTGTACCGCGGAGTACAGGATCCATGTCCGCCAGCCCGCGGAATCC 195
Db	1223 CACACCACTTCACGCGCGGCGAGCGGAGAACACGAGTTTCGCGCGCGCGCGGAGATCG 1164
QY	196 TGCTCTCTTCGCATGATCTCGTTCGCGCGTTCACGCGGAGGACGAGCGGTAGTCCACC 255
Db	1163 CGTTCGCGGCGGACCATCGGTTCGCGGTAGATCCAGGAGGACCGAGACCGGTTCGGC 1104
QY	256 GCGTCGCGGCGTGAACCGTTCGCGGCTCGCACCGGATGTGCGTCCGCGGCGGTGAGCGG 315
Db	1103 TCAGCGCGCGCGGCTCTCTTCGAGACGATCGGATGTTCGTCGCGCGGCGGTGAGCAG 1044



<http://www.nih.gov/jp-jun/cgi-bin/frameplot.pl>. Where possible we chose an initiation codon (atg, gtg, ctg or tgg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13 bp before the initiation codon). If this could not be identified we chose the most upstream initiation codon.

Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes (three PKS gene clusters for lankacidin, lankamycin and an unknown type II polyketide, a carotenoid biosynthetic gene cluster, many regulatory genes and others). The range of each biosynthetic gene cluster has been deduced by comparison with similar gene clusters in most cases and may be revised in future.

## FEATURES

## source

Location/Qualifiers

1..210614

/organism="Streptomyces rochei"

/mol\_type="genomic DNA"

/strain="7434AM4"

/db\_xref="taxon:1928"

/plasmid="pSLA2-L"

/notes="linear plasmid"

1..1992

/note="left terminal inverted repeat, TIR-L; shows 99.4 % (1981/1992) sequence identity to TIR-R (complement (208623..210614))"

683..2188

/note="N-terminal sequence is almost identical (435/437) with that of ORF143 at the right end of pSLA2-L until the inner end of TIR-L"

ORF1 (501 aa)

similar to AL590463 Streptomyces coelicolor putative helicase, SCPI.136 (879 aa); homology is seen until the inner end of TIR-L"

/codon\_start=1

/transl\_table=11

/product="putative helicase"

/protein\_id="BAC76459.1"

/db\_xref="GI:30698346"

/translation="MTGTGKTQDREAAQOEADVAVRLELPVLLIAPRGLRTQVIM ATGSGTRVAARSAEKLRAGRVLVPSLDLITQTEAAWEAGRTGPMIGVSLRGED VAPNTDTEELVDVYRPEKVTYFATYASLGILTEARHGGELPCMDLIVYDEAHT SGRGLFWAVHNDNTRIPSLRLNMTATFRLWQLEDACAGSELVAMSEDDPDGLFG ARCTFLTSEADRGICAPYVVCVDITDITQQAQLLGVGRSDEVRGARLAALQTA LKASSEENFRFTLVPHVMKEAFAGLPDVAKLHAAGLGLYPRITWANNLCGEH KPGRHRRVFTGIGATDGTGVVEKGLSVKVLGEGVTRCDYVWADVSGMPDL VQAVGRALRTQPGQKVASLWVPLDPCGETADNMLTSRPNGLALLPQAPLLTGHG EGFRAARPGREVRVPLVGRGAQRGASGPASRTTPSAMSLSLSNRP"

complement (2315..3595)

/note="ORF2 (426 aa)

similar to AE004736-10 Pseudomonas aeruginosa hypothetical protein (442 aa)

/codon\_start=1

/transl\_table=11

/product="putative secreted protein"

/protein\_id="BAC76460.1"

/db\_xref="GI:30698347"

/translation="MRKKPALLGIPGLLVAVLAPAPAAHASADTYVTWAPPTAA PADATVYEDQTLRQVHLSVAGDSLRVTFNEFGSPITIGEVAARPAAGQATPAV DQGTDRVVRGGPSPATLAFGTQWSDPVALTAGDGLVLSLPTQPGTIVSHAA YQNFVAAGVTCAPDLTPVSTATSWHFLSGVSDRRGADSVSVTLGDSITDGEHT TLADRWPPDLLAERLRDGLAGTGVWAGIGGNRLLRDPDPGSGASPAEPAAYFGE SALKRDPDRVQPGAPAVTVLLGVNDLGGPGIAAPASDEVTAELIACYROLIERAH EHLKLYGALITPAGTIGYFTFPRREARQQVNDWIRISGAFDVLDPDAVLDPAR PDHLLPAYDGGDLHNDAGNAWAFPLDSLR"

complement (3659..4645)

/note="ORF3 (328 aa)

similar to Y00459-2 Streptomyces griseus regulatory protein, StrR (350 aa)

/codon\_start=1

/transl\_table=11

/product="StrR-like regulatory protein"

/protein\_id="BAC76461.1"

/db\_xref="GI:30698348"

/translation="MTAALAGNLDQETRSIPISLVPGYSPRLCGDAEHVARLADTD EALPILVERSTLRVIDGHRVLAAKAKGHTTIEVALFDGAEEAFLLAVRSNMTHGL

CDS

CDS

PLSRQDRRAAQRILAQWPHLSDRVAGIAGAKTVAALRPLAAGTFNPOARRGRD GRPLDGTGRKAAELIAOPQASVREVARHAGISPAVASDVRRLRAGSPVPER NTCAREAGTSGSATPGGAAPEPVSVVRPIRPEDSPVLRLLRDPSLRHKSGRRL LRLQCGAVERTLAMAQVPPHCITDLVAELAREYADLWAEFAREVARTG"

complement (4322..5812)

/note="ORF4 (296 aa), lankacidin biosynthesis protein similar to AE004625-4 Pseudomonas aeruginosa pyrroloquinoline quinone biosynthesis protein B (304 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein B"

/protein\_id="BAC76462.1"

/db\_xref="GI:30698349"

/translation="MILLGTAAGGPPQWNCACALCARGRRGELPARSQCVASVGDG RDWLLNSPDITQLLAPALTFGPGPRTFVRGVLITDAEVDHGLAVLRGATGL TVYAAPVRGALSALFVRGLLDYAPWRDATAPGFAAGGLTVTAHPVGTAPK YAHAPDPAPVCAYRIEDPATGALVYAPLATPDPGDDLLASATCALLDGTFFSA GELGTATSSAGAGSLMGHLPVAGPGGSLAALHRGLRRIYTHLANTNLLDPSAA HAAVREAGVEVLDGSELVL"

complement (5815..6909)

/note="ORF5 (364 aa), lankacidin biosynthesis protein similar to AJ277117-5 Gluconobacter oxydans putative pqqe protein (359 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein E"

/protein\_id="BAC76463.1"

/db\_xref="GI:30698350"

/translation="MADPAVGAPAGMLIETHRCPLHCPYCSNPDLVLRREALTCQ WTDILTQARELVGVVQHFSGEPLAPDLPLDVGARLRGAYVNLVTSVGLTAEARH DLARGVDHVQLSDADPAAGQAIAGARVHTAKLEAARVTAAGLPLTVNIVLHRGN IDRTGRVLDIADLDRIELANTQYVWGLRNRAALMTAAQALAAAREAVRHARTY AGGBELVYAADYVDPKPKMDGSGTQLTVPAGDVLPCPAAYATTTLPVENALRR PLSEIWTASRSFNAVYRTGMRPEFCRCPERHADHGGCRQCAQLTGDAAATDPACGL SPHSLDAALAEVTDGVPFAFVRGPVPA"

complement (6902..7174)

/note="ORF6 (90 aa), lankacidin biosynthesis protein similar to AL603642-200 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein D (98 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein D"

/protein\_id="BAC76464.1"

/db\_xref="GI:30698351"

/translation="MTGLPEPTVPRLPVRLTRDPARGELALLPERVWVLDATAAV LAHCDGTTSLAGIVERLAEEVGSVEDVRELLLRQAQRVVDLHG"

complement (7171..7899)

/note="ORF7 (242 aa), lankacidin biosynthesis protein similar to AE004625-5 Pseudomonas aeruginosa pyrroloquinoline quinone biosynthesis protein C (250 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein C"

/protein\_id="BAC76465.1"

/db\_xref="GI:30698352"

/translation="MSMSVREVAAPSEAFRORLHALESSYNDRHPFRHMEGLL DEGLRLWAANRYQCLPQKDAIVANCLPVRQWLSRVYHDCGACAGAEK WLRLEAVRWYQDHDLRLVACTRFAVDYVDFARRPWLAAAGSLTFLPFGPLL AHRGLRHPYFWIAEEGFEYFTARIIEVVGSGRSLLDLVARHAVSRQEQACVRALA FKRCRLNVLSDIYHTGNGATRS"

complement (7926..8030)

/note="ORF8 (34 aa), lankacidin biosynthesis protein similar to AL603642-197 Sinorhizobium meliloti putative pyrroloquinoline quinone biosynthesis protein A (31 aa)"

/codon\_start=1

/transl\_table=11

/product="pyrroloquinoline quinone biosynthesis protein A"

/protein\_id="BAC76466.1"

/db\_xref="GI:30698353"

/translation="WRTSGKGLPAKAWHRPDPFTVITDTCMEVTAFFSR"

8420..10057

/note="ORF9 (545 aa), possible lankacidin resistance protein

CDS

CDS



```

Db      45097  TGTGCGAGCCCTGTCCCGGAGCGGTGAGGAATTCCAGAGTCCGTCACGGGACTCCT 45156
Oy      482   CGACGTGCGGCGGAAC 498
Db      45157  CGTCCGCGGCGGAAC 45173

RESULT 43
AF333038/c
LOCUS   AF333038
DEFINITION Streptomyces viridochromogenes Avilamycin A biosynthetic gene cluster, complete sequence.
ACCESSION AF333038
VERSION   AF333038.2
KEYWORDS  GI:28380396
SOURCE   Streptomyces viridochromogenes
ORGANISM Streptomyces viridochromogenes
REFERENCE 1 (bases 1 to 59816)
AUTHORS   Decker, H., Gaisner, S., Pelzer, S., Schneider, P., Westrich, L., Wohleben, W. and Bechtold, A.
TITLE     A general approach for cloning and characterizing dNDP-glucose dehydratase genes from actinomycetes
JOURNAL   FEMS Microbiol. Lett. 141 (2-3), 195-201 (1996)
MEDLINE   96313316
PUBMED    8768522
REFERENCE 2 (bases 17444 to 23885)
AUTHORS   Gaisner, S., Trefzer, A., Stockert, S., Kirschning, A. and Bechtold, A.
TITLE     Cloning of an avilamycin biosynthetic gene cluster from Streptomyces viridochromogenes Tu57
JOURNAL   J. Bacteriol. 179 (20), 6271-6278 (1997)
MEDLINE   97474248
PUBMED    9335272
REFERENCE 3 (bases 4931 to 5794; 6930 to 7681; 34188 to 35953)
AUTHORS   Weitnauer, G., Gaisner, S., Trefzer, A., Stockert, S., Westrich, L., Quiros, L.M., Mendes, C., Salas, J.A. and Bechtold, A.
TITLE     An ATP-binding cassette transporter and two RNA methyltransferases are involved in resistance to avilamycin in the producer organism Streptomyces viridochromogenes Tu57
JOURNAL   Antimicrob. Agents Chemother. 45 (3), 690-695 (2001)
MEDLINE   21091857
PUBMED    11181344
REFERENCE 4 (bases 1 to 59816)
AUTHORS   Weitnauer, G., Muehlenweg, A., Trefzer, A., Hoffmeister, D., Sussmuth, R.D., Jung, G., Welzel, K., Vente, A., Girreser, U. and Bechtold, A.
TITLE     Biosynthesis of the orthosomycin antibiotic avilamycin A: deductions from the molecular analysis of the avl biosynthetic gene cluster of Streptomyces viridochromogenes Tu57 and production of new antibiotics
JOURNAL   Chem. Biol. 8 (6), 569-581 (2001)
MEDLINE   21303144
PUBMED    11410376
REFERENCE 5 (bases 22882 to 23361)
AUTHORS   Bechtold, A.
TITLE     Direct Submission
JOURNAL   Submitted (24-MAY-1996) Universitaet Tuebingen, Pharmazeutisches Institut, Pharmazeutische Biologie, Auf der Morgenstelle 28, Tuebingen D 72076, FRG
REFERENCE 6 (bases 17444 to 23885)
AUTHORS   Bechtold, A.
TITLE     Direct Submission
JOURNAL   Submitted (20-MAR-1997) Universitaet Tuebingen, Pharmazeutisches Institut, Pharmazeutische Biologie, Auf der Morgenstelle 28, Tuebingen D 72076, FRG
REFERENCE 7 (bases 4931 to 5794; 6930 to 7681; 34188 to 35953)
AUTHORS   Gaisner, S. and Bechtold, A.
TITLE     Direct Submission
JOURNAL   Submitted (01-NOV-2000) Pharmazeutische Biologie, University of Kiel, Gutenbergstrasse 76, Kiel 24118, Germany
REFERENCE 8 (bases 1 to 59816)

```

---

```

AUTHORS   Weitnauer, G., Trefzer, A., Muehlenweg, A., Hoffmeister, D. and Bechtold, A.
TITLE     Direct Submission
JOURNAL   Submitted (02-JAN-2001) Pharmazeutische Biologie, Christian Albrechts Universitaet, Gutenbergstrasse 76, Kiel 24118, Germany
REFERENCE 9 (bases 1 to 59816)
AUTHORS   Mosbacher, T., Weitnauer, G., Bechtold, A. and Schulz, G.E.
TITLE     Direct Submission
JOURNAL   Submitted (13-DEC-2002) Institut fur Pharmazeutische Biologie, Universitaet Freiburg, Stefan-Meier-Str. 19, Freiburg 79104, Germany
REMARK    Sequence update by submitter
COMMENT   On Feb 14, 2003 this sequence version replaced gi:15077433.
FEATURES
     source
         1..59816
             /organism="Streptomyces viridochromogenes"
             /mol_type="genomic DNA"
             /strain="Tues57"
             /db_xref="taxon:1938"
             /gene="avlX1"
             /CDS
                 1590..2678
                     /gene="avlX1"
                     /note="avlX1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative GTP binding protein"
                     /protein_id="AAK83161.1"
                     /db_xref="GI:15077434"
                     /translation="MSLTIGVIGLPNVGKSTLENALTKNDVLAANYPFATIEPNVGV
                     GVPARITKLAIEFFSSQRIIPATVDFDIAGIVRGASEGGLGNKFLANIRESDAICQ
                     VIRAFKDNVHVVDGKVPKDDIETINTEILGLADLQTIKRVLPRLQKSRKKDLAPK
                     KVAYEAKIEILRGDTLFAHGIQVGTETNELLDHLLTKPFLYFNVNDELELID
                     FKNEQALVAPAEAFILNAKLEADLAELDDLELLOQSGQDEPGLATLARVGNTL
                     GLQYLTAGPKESRAWTIKKGATAPEAGVHTDFQKGFKAIEVISFTDLVETGSAE
                     ARAKGKARMEGKYVMQGDVVEFRNV"
                 2870..3151
                     /gene="avlX2"
                     /CDS
                         2870..3151
                             /gene="avlX2"
                             /codon_start=1
                             /transl_table=11
                             /product="avlX2"
                             /protein_id="AAK83162.1"
                             /db_xref="GI:15077435"
                             /translation="MSINRERSKALFPLIKKNDNEAEIIVSKHGNVLVSAEDYA
                             ALREGSYLLRSPANARLLKAYENALAHVNVSEKLIIDPDSDAGSGAA"
                 3148..3402
                     /gene="avlX3"
                     /CDS
                         3148..3402
                             /gene="avlX3"
                             /codon_start=1
                             /transl_table=11
                             /product="avlX3"
                             /protein_id="AAK83163.1"
                             /db_xref="GI:15077436"
                             /translation="MRLVFDQGMDDYTSWLKNDKMLARINKLIEDVRDPFTGICK
                             PEPLKYLPGAWSRRIDDEHRLVYLVTDKEIVILAAARYH"
                 3410..4310
                     /gene="avlX4"
                     /CDS
                         3410..4310
                             /gene="avlX4"
                             /codon_start=1
                             /transl_table=11
                             /product="avlX4"
                             /protein_id="AAK83164.1"
                             /db_xref="GI:15077437"
                             /translation="MILHSRFRWVADRTYGISGTGLDWLDSAPNHLVVEHSHTWS
                             LLEASAPVGSIFVATWIDRTLYPERMPFPQHSBESGYGVYIYVTSKSDSRP
                             RCRDLKSGRVLRLPGWPHAEFRHLIDDAATITHAEWEALIGYWRSLTAAMLIGYV
                             RRATFRERIGLPPASFCFSGAHCFTLARFGTHADAEILTGTDROMIPSNWSALAL
                             WTGPYCPIGR"
                     complement(4307..4952)
                         /gene="avlX5"

```

```
CDS
complement(4307..4852)
/gene="aviX5"
/codon_start=1
/transl_table=11
/product="aviX5"
/protein_id="AAK83165.1"
/db_xref="GI:15077438"
/translation="MTSHGINAPCAQHGRYVYRPPAEDEPLDPAMRPVLTTRAPPEDR
AADRAWPDPDRPALCWLPTSGVMATRRTTTTPHIETLVQQLDGDSDASVAAQAQLVD
IGRDAITLIGGLPTLGSFGLTAIEVEFVADPRCCPALIALLDADDPVREWARGP
DRTERPPTSRRSSTTWTTTK"
4931..5794
/gene="aviRb"
4931..5794
/gene="aviRb"
/feature="AviRb"
/codon_start=1
/transl_table=11
/product="rRNA methyltransferase"
/protein_id="AAG32066.1"
/db_xref="GI:11141830"
/translation="MAKRGERTPAARRITRNARFQQWALLGNRNKTRAGEFILVM
GVRFISLAVHGMVPTLLYDQRELSKWARLLRTVTEQIAMADLLMELGEKNEA
PPEVAVVEMPADLDLPVREDPLVLPDRPTSPGNIGSIIRSDALGAHGLIVAGH
AADYDPKSVRSSTGSLFSLPAVRVPSPGEVMDVWEARRAAGTPIVLGTDHGDGV
FDPTOPTULLIGNETAGLSNAWRILCDYTVSIPWAGSASSLNAANAATAILYEVR
QRISGRATTP"
5809..6159
/gene="aviX6"
5809..6159
/gene="aviX6"
/codon_start=1
/transl_table=11
/product="aviX6"
/protein_id="AAK83166.1"
/db_xref="GI:15077439"
/translation="MYLARASWIRPTISACRSKSPARRPSVAAGSEPSVRASTSPSY
TDMNSAGSGTGREPDASVSTRANSPVSATSGLTSTGTMTARPTSVTTCNPNGTG
WTSPTLPQFRLPRR"
6159..6305
/gene="aviX7"
6159..6305
/gene="aviX7"
/codon_start=1
/transl_table=11
/product="aviX7"
/protein_id="AAK83167.1"
/db_xref="GI:15077440"
/translation="MSSKPGQVSPSHRPQTRIRSRACLAASTTSMNGFHRPQLLPM
YCS"
6489..6830
/gene="aviX8"
6489..6830
/gene="aviX8"
/codon_start=1
/transl_table=11
/product="aviX8"
/protein_id="AAK83168.1"
/db_xref="GI:15077441"
/translation="MATPPKWPSEIDYQAPVPGKRAQGFOLLVRIQLPGLHEHKMCS
PAHVLRLKMPRDLVHPRKRAADPLTARFTRLHIGNGLAESRLPMKIPGVHPG
FRHSLPRTH"
complement(6930..7682)
/gene="aviRa"

Query Match      16.2%; Score 81; DB 1; Length 59816;
Best Local Similarity 50.9%; Pred. No. 0.038;
Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGGTCAGTCGTACGTCACGCGGGACGTACAGGATCCACTGTCGCGCAG 181
DB 15020 CTCGCCACCCCGATCACACAATTCACGCGTGGCGGGAAGACACAGGCTGCTCGG 14961

QY 182 CCGCGCGGAACCTCTGTCTCTTCGCCATGATCTCTGTGGGCGTGTTCAGGCGAAGACCA 241
DB 14960 AGGCAGGAAGTCGCGTTCCCGGGCGACCATCGCGTCCCGGTAGATCCACGGGAGGACCA 14901

QY 242 GCGCGTAGTCCACCGGTCGCGGCGTGAACGGTCCGGGTCGCGACCGGGATGTCGTCG 301
DB 14900 GGAACGTGTGGGCTTAAGCCCGCGGCTCTCTCTCGAGACGATCGGATGTCGTCG 14841

QY 302 CCGGGGTGAGCCGCGCTCTGTTCGCGCGCTGTGTGCGACACCCAGGAGACCAAGGTCG 361
DB 14840 CCGGCGTGTAGCAGCGCAACTTGTCTCGTTCACTCGCGATGTCAGGCGAGGCGTCT 14781

QY 362 GACCGATGCGCGAGAGTTCGTACGGTGGCGCTCTTCGCCGTGCGCGCTACGCCACCA 421
DB 14780 CGTCAGCGCGCAGTACTGCAGATGACGTTGCCCTTCTGTGAGCGCGCTAGGCCAAGG 14721

QY 422 CCGGCTTCCCTCGGCTTGAGCGAGTTCAGCAGGCGAGAGTTCGCTGCGGATGCGCT 481
DB 14720 TGTGAGCGCTTGTTCGCGCGGCGGTGAGGAATTCAGGAGTCTGTTCAGGAGTCTCT 14661

QY 482 CGACGTCGCGCGCAAC 498
DB 14660 CGGTCGCGCGCGCAAC 14644

RESULT 44
AX535308/c AX535308 1224 bp DNA linear PAT 22-NOV-2002
LOCUS Sequence 45 from Patent WO02068436.
DEFINITION AX535308
ACCESSION AX535308
VERSION AX535308.1 GI:25261950
KEYWORDS
SOURCE Streptomyces viridochromogenes
ORGANISM Streptomyces viridochromogenes
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1
AUTHORS Weinbauer,G., Muehlenweg,A., Trefzer,A. and Bechthold,A.
TITLE Avilamycin derivatives
JOURNAL Patent: WO 02068436-A 45 06-SBP-2002;
Combinature Biopharm AG (DE)
FEATURES
Location/Qualifiers
source 1..1224
/organism="Streptomyces viridochromogenes"
/mol_type="unassigned DNA"
/db_xref="taxon:1938"
/feature="Teillesequenz avigidna des codierenden Stranges 1."
Das Nucleotid 1 entspricht dem Nucleotid 13784 des
codierenden Stranges 1."

ORIGIN
Query Match      16.2%; Score 80.8; DB 6; Length 1224;
Best Local Similarity 53.1%; Pred. No. 0.12;
Matches 172; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 175 CCGCCAGCCCGGGAACCTCTTCGCCATGATCTCGTCGCGGTGTTCCAGGCG 234
DB 1184 CTTCCGAGCGCAGAGTCTCGTCCCGGGACACNTCGCTCCGGTAGATCCACGG 1125

QY 235 AAGAGCAGCGGTAGTTCACCGCTGCGCGCTGAACCGGTCCGGGTGCGCACCGGATG 294
DB 1124 AGGACCGAAGTCTGGTCGGGCTCAAGCGCCCGGGCTCTCTCTCGAGACGATCGGATG 1065

QY 295 TCGTCCCGGGGTGAGCGGCTCTGTCGCGCGGCTGTCGCACACCCAGGAGACC 354
DB 1064 TTGTCGCGCGGGGTGTAGCAGCGCACTTGTCTGTTCACTTCGCGGTGAGGCGGAGG 1005

QY 355 AGTCCGCGACCGATGCGCGAGAGTTCGTACGCGTGGCGCTCTTCGCGCGCGGTAC 414
DB 1004 AGCGTCTCTCCAGGCGCGAGTACTGCAGATGACGTTGCCCTTCTGTTCAGGCGCGGTAG 945

QY 415 GCACACCGCGCTTGCCTTCGCGCTTACGAGGATTCAGCAGGCGCGAGGTCGCTGCGG 474
```

```

Db      944  CCAAGGGTGTGACGCCCTTGTCCCGGAGCGGTGAGGAATCCAGGATCGGTACGG 885
Qy      475  ATGCCCTCGACGTCCGGCGGCAAC 498
Db      884  GACTCTCGGTCCCGCGGCGAAC 861

RESULT 45
LOCUS   AE009716
DEFINITION Brucella melitensis 16M DNA linear BCT 20-MAR-2003
           complete sequence.
ACCESSION AE009716 AE008918
VERSION   AE009716.1 GI:17985017
KEYWORDS
SOURCE
ORGANISM Brucella melitensis 16M
           Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
           Brucellaceae; Brucella.
REFERENCE 1 (bases 1 to 11659)
AUTHORS DelVecchio,V.G., Kapural,V., Redkar,R.J., Patra,G., Mujer,C.,
           Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A., Lykidis,A.,
           Reznik,G., Jablonski,L., Larsen,N., D'Souza,M., Bernal,A.,
           Mazur,M., Goltsman,E., Selkov,E., Elzer,P.H., Hagius,S.,
           O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kyrpides,N. and
           Overbeek,R.
TITLE     The genome sequence of the facultative intracellular pathogen
JOURNAL   Brucella melitensis
PUBMED    Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
REFERENCE 11756688
AUTHORS   DelVecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.
TITLE     Direct Submission
JOURNAL   Submitted (13-NOV-2001) Institute of Molecular Biology and
           Medicine, University of Scranton, Scranton, PA 18510, USA
REFERENCE 3 (bases 1 to 11659)
AUTHORS   Elzer,P.H. and Hagius,S.
TITLE     Direct Submission
JOURNAL   Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag
           Center, 111 Dairymple Building, Baton Rouge, LA 70803, USA
REFERENCE 4 (bases 1 to 11659)
AUTHORS   Kapural,V., Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A.,
           Lykidis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M.,
           Bernal,A., Mazur,M., Goltsman,E., Selkov,E., Haselkorn,R.,
           Kyrpides,N. and Overbeek,R.
TITLE     Direct Submission
JOURNAL   Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
           Park Drive, IL 60612, USA
REFERENCE 5 (bases 1 to 11659)
AUTHORS   Letesson,J.-J.
TITLE     Direct Submission
JOURNAL   Submitted (13-NOV-2001) Unite de Recherche en Biologie Molculaire,
           Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,
           61 rue de Bruxelles, Namur 5000, Belgium
REFERENCE 6 (bases 1 to 11659)
AUTHORS   O'Callaghan,D.
TITLE     Direct Submission
JOURNAL   Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue
           Kennedy, Nines 30900, France
FEATURES
         source
         1..11659
            /organism="Brucella melitensis 16M"
            /mol_type="genomic DNA"
            /strain="16M"
            /db_xref="taxon:224914"
            /chromosome="II"
            complement(73..972)
            /gene="BMEII0820"
            complement(73..972)
            /gene="BMEII0820"
            /codon_start=1
            /transl_table=11
            /product="ALS OPERON REGULATORY PROTEIN"

gene
CDS

```

```

/protein_id="AAL54062.1"
/db_xref="GI:17985018"
/translation="MDIPMELRHLRYFLAVAEENFTRAAKLIGIQPPLSQIRDL
NEIGVLFHRVPHGAELTAAGAEFLAEETATIASSEKAKLAQAARGESGRLSIGFT
ASSAFNPISGAIRHPRAKWPEVLEFLTEMNTNLLIERLVGELDTAIFRPLEDPA
VMKRLADEPMLVALPSHRLAVYERLPLSALAGEPFLVFRPIVGLSLVDYDAACRM
AGFENVVTEAPQPMPSVNVLAANLGSIVPAATAQIRLDGVTYRPIEGEPVVARLAL
ASMSKLSRSPVDNLLISLPRBCI"
1065..1778
/gene="BMEII0821"
CDS
1065..1778
/gene="BMEII0821"
/codon_start=1
/transl_table=11
/product="MULTIDRUG RESISTANCE PROTEIN"
/protein_id="AAL54063.1"
/db_xref="GI:17985019"
/translation="MSRIEDRSNGIAAADNGLTLVEMPEBAKTYLTGMPAYRNASL
ALFLSGFATFSLYCVQPLMPLFAKDFSTIPTASSLSLSLSTGFLAFAIPFAAIAER
FQRKSLPASLIGSALCTIGCAVVPDWHMLLARSLOGLLGGVPAAVMSVLAIEIDP
RGLGASMLGIYAGNAGMGAGRVITGTLAEYYSWRPALAVMGLLGLAAAIQFWALLPT
SRNTPFGMEPATYRPG"
1643..2365
/gene="BMEII0822"
CDS
1643..2365
/gene="BMEII0822"
/codon_start=1
/transl_table=11
/product="MULTIDRUG RESISTANCE PROTEIN"
/protein_id="AAL54064.1"
/db_xref="GI:17985020"
/translation="MAPSPCRHGARSNGSSNRLGIVAYVQKFAQNRMGARFHVSAW
LSHLNRPALPMLIAIGFTLVMGPFVTVYINIGFLRIEPPNLSLSTGFLAFAIPFAAIAER
ASWAGRIGDRGLGHTMPLIALAVAAAGCLITLASLPIIVGGIIFLTGDFMGSHVA
SALVRLALBIKHASSILYLAAYLGSSIAGSAGGYFWTADGWAVADFTFAMLALAP
AALAAAWFAPEGQTDKPLTF"
2492..2659
/gene="BMEII0823"
CDS
2492..2659
/gene="BMEII0823"
/EC_number="2.7.1.30"
/codon_start=1
/transl_table=11
/product="GLYCEROL KINASE"
/protein_id="AAL54065.1"
/db_xref="GI:17985021"
/translation="MSGYILAIDQGTSTRMLFDRNRRVVVGLQCEFTQHPSSGWV
BHDAAEIWKSV"
2918..3988
/gene="BMEII0824"
CDS
2918..3988
/gene="BMEII0824"
/EC_number="2.7.1.30"
/codon_start=1
/transl_table=11
/product="GLYCEROL KINASE"
/protein_id="AAL54066.1"
/db_xref="GI:17985022"
/translation="MLNRPGLRERRAKQVCFTIDSWLIYKLTGCAKHWDTATNAS
RTLIYHIGENRWDDLLDILGIPAAALPEVKDCAADFGMTDFPALFGVSIPTLVAGDQ
QAIVGNACPEPGMKSTYGTCCFALNTGTDRVTSNRLITTYATLDGTYTVALEG
SAVITGAAVOWLRDEMGPFISVASEVSALAEKADPNQRIYLVPAFTGLGAPYWDAAERG
AIFGLRTGTGRAAFARALBSVAYOTFDLLEAMQGDWKGATNHTVLRVDGGMVADWT
MORLADILINAPVDRPVLETTVLGAAMLAASRAGIWPDRKGFSPERQDRCFEPAMPE
KERESATAGWDSVSRCLTRPQ"
4105..5385
/gene="BMEII0825"
CDS
4105..5385
/gene="BMEII0825"
/codon_start=1
/transl_table=11
/product="URACIL PERMEASE"
/protein_id="AAL54067.1"

```

	Query Match	15.6%	Score 78	DB 1	Length 11659
	Best Local Similarity	51.1%	Pred. No. 0.15		
	Matches 183	Conservative 0	Mismatches 175	Indels 0	Gaps 0
Qy	140	TCAGTGTACGCGGGACGTACAGGATCCACTCTCCGCCAGCGCGCGGAATCTCTGCT	199		
Db	7078	TCAGGTGACATGGGAATGGGAATGAATTTGCGCGCCCATTCACGACATGCTGCA	7137		
Qy	200	CCTTCGCCCATGATCTCTCGCGCTGGTTTCAGGCGAAGACAGCGCGTACTCCACGCGCT	259		



```
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN33634.1"

Query Match      15.6%; Score 78; DB 1; Length 11852;
Best Local Similarity 51.1%; Pred. No. 0.14;
Matches 183; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Qy 140 TCACGTGTACCGGGGAGGACGATCCAGTCCGCGCAGCGCGCGGAACTCTGCT 199
Db 8791 TCAAGGTGACATGGGGAATGGGAATGATGAATTCGCGCGCCCATTCACGACATGCTGCA 8732

Qy 200 CTTTCGCGATGATCTCGTCGCGCGGTTCACAGGGAAGAGCAGCGGTAGTCACCGCGT 259
Db 8731 TCTCCCGATATCTCTGCTTGAATTCATGCGCAGATGAGATGATGTCGGCGGT 8672

Qy 260 CGGCGGTGAACCGTCCGCGGTGCGCACCGGATGTGCGTCCGCGGGGTGAGCGCGCCT 319
Db 8671 AATGTGATCGCGCAACATCGTAATCGAATGTGCAITCCAGGCGTAAAGCGCGCT 8612

Qy 320 GCTTGGCGCGGTGTCGTCGACACCGCAGGACAGGTCCGACCGATCCGCGCAGAGT 379
Db 8611 GCTTGTAGGATTCGATTCACGCTGAAACCCAGAAATCTGTCGCCATGCCGATAGT 8552

Qy 380 TCGTCACGCTGCGCTCTTCGCGCGGTGCGCGGTGCGCACCGCTTCTGCTCGCGCT 439
Db 8551 TCAAAAGCGTGTGCTTCTGCGCGGTGCGCGGTGCGCACAGATGCTTTCGCGCATTT 8492

Qy 440 TGAGCGAGTTCAGCAGGCGGAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 497
Db 8491 TGGCGCAATCAGGAAGACAGAGATCACGCTTGTGCGCGGTCTTTTCGCGCAA 8434

RESULT 47
AF076290
LOCUS
DEFINITION
AF076290 36502 bp DNA linear BCT 05-NOV-2001
Brucella melitensis Bme18 (bme18) gene, partial cds; Bme19 (bme19),
Bme16 (bme16), Bme17 (bme17), Bme20 (bme20), Bme21 (bme21), Bme22
(bme22), Bme23 (bme23), Bme24 (bme24), Bme25 (bme25), Bme26
(bme26), Bme27 (bme27), Bme28 (bme28), Bme29 (bme29), Bme30
(bme30), Bme31 (bme31), Bme1 (bme1), Bme2 (bme2), Bme3 (bme3), Bme4
(bme4), outer membrane protein (omp31), Bme6 (bme6), Bme7 (bme7),
Bme8 (bme8), Bme9 (bme9), Bme10 (bme10), Bme11 (bme11), Bme12
(bme12), Bme13 (bme13), and Bme14 (bme14) genes, complete cds; and
Bme15 (bme15) gene, partial cds.
AF076290 AF076289 U39453
AF076290.2 GI:16740531

Brucella melitensis
Brucella melitensis
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Brucellaceae; Brucella.
1 (bases 22692 to 23728)
Vizcaino, N., Cloeckaert, A., Zygmunt, M.S. and Dubray, G.
Cloning, nucleotide sequence, and expression of the Brucella
melitensis omp31 gene coding for an immunogenic major outer
membrane protein
Infect. Immun. 64 (9), 3744-3751 (1996)
6355886
8751924
2 (bases 1 to 36502)
Vizcaino, N., Verger, J.M., Grayon, M., Zygmunt, M.S. and Cloeckaert, A.
DNA polymorphism at the omp-31 locus of Brucella spp.: evidence for
a large deletion in Brucella abortus, and other species-specific
markers
Microbiology 143 (Pt 9), 2913-2921 (1997)
97453476
9308175
3 (bases 1 to 36502)
Vizcaino, N., Cloeckaert, A., Zygmunt, M.S. and Fernandez-Lago, L.
Molecular characterization of a Brucella species large DNA fragment
deleted in Brucella abortus strains: evidence for a locus involved
in the synthesis of a polysaccharide
```

```
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
Infect. Immun. 67 (6), 2700-2712 (1999)
99270925
10338472
4 (bases 1 to 36502)
Vizcaino, N., Cloeckaert, A., Zygmunt, M.S. and Fernandez-Lago, L.
Characterization of a Brucella species 25-kilobase DNA fragment
deleted from Brucella abortus reveals a large gene cluster related
to the synthesis of a polysaccharide
Infect. Immun. 69 (11), 6738-6748 (2001)
21481949
11598046
5 (bases 1 to 36502)
Vizcaino, N.
Direct Submission
Submitted (01-JUL-1998) Microbiologia y Genetica, Universidad de
Salamanca, Edificio Departamental. Avda. Campo Charro s/n,
Salamanca 37007, Spain
6 (bases 1 to 36502)
Vizcaino, N.
Direct Submission
Submitted (31-MAY-2001) Microbiologia y Genetica, Universidad de
Salamanca, Edificio Departamental. Avda. Campo Charro s/n,
Salamanca 37007, Spain
Sequence update by submitter
On or before Nov 5, 2001 this sequence version replaced gi:4927378,
gi:5478231.
FEATURES
Location/Qualifiers
1..36502
/organism="Brucella melitensis"
/mol_type="genomic DNA"
/strain="16M"
/db_xref="taxon:29459"
<1..459
/gene="bme18"
<1..459
/gene="bme18"
/notes="putative transport protein"
/codon_start=1
/transl_table=11
/product="Bme18"
/protein_id="AAD43832.2"
/db_xref="GI:16740532"
/translation="ELGIFTVYLFVFGIAASWSAGRIGDRLGHFTMLPIALAVEAGCL
LTLSASLPVIGGIIFLVGFMSHSVASALVGRALRIKGHASLSLYLLAVYLGSSIA
GSAGYFTWADGNWAVDFTFAMLAALAAALAAWFAREQGTDKPGTTF"
586..2082
/notes="similar to glycerol kinases found in other
bacterial species; contains stop codon"
misc_difference
751..753
/notes="stop codon detected in the reference strains of the
three Brucella melitensis biovars and in the
Brucella abortus reference strain, but not in the other
Brucella species reference strains"
970..2082
/gene="bme19"
970..2082
/gene="bme19"
/notes="putative glycerol kinase"
/codon_start=1
/transl_table=11
/product="Bme19"
/protein_id="AAL27667.1"
/db_xref="GI:16740533"
/translation="MLLDPYFSGTKLAWLHNPGLRBRRAQKQVCFGTIDSLIYKL
TGKARVDTNASTLTLYHIGENWDELLDILGIPAAMLPEVKOCADFGMTDPAL
FGVSPILGVAGDQQAATVGNACPEFGMKSTYGCFCALLNTGTDVTVSSNRLTTI
ATRLGVTYALGSGIFTAGAAVQWRDEMFGISVASVSALEKADPNORLYVPAP
TGLGAPYWDAAEARGAIFGTRGTGRAEFAALESVAVQTFDLLEMQDQKGATNHT
VLRVDGQWVASDWTQRLADILNAPVDFVLETTVLGAAWLAASRAGIWPDKRGFSE
RNQRCRPEPAMPEKERSAIAWSDSVSRCLTRPQ"
2199..3479
/gene="bme16"
2199..3479
/gene="bme16"
misc_feature
misc_difference
gene
CDS
```



/gene="bme16"  
/note="putative uracil permease"  
/codon\_start=1  
/transl\_table=11  
/product="Bme16"  
/protein\_id="AAD33089.2"  
/db\_xref="GI:16740538"  
/translation="MGYFSPRLTNSAVLPDERLPAAAVPMGIQHLAMSGSTIVA  
PILMGDPNVAFFSGIGILLFLITAGRPVPSYLGSSFAIVAVIAATGYSGGPNP  
IAIALGGIACGLAISLIGLIVNFGTGWKELMPFAVGAIGVIGLNLAPVAIGQL  
KGTGAHTSIALTILVCWAMISAYPGEGTRRLSVLPAVGLYLLIAGNGLVIFGID  
FVAKNAWAFGPHFTPAFNQWAVLIAVAIVLVAENLGHKALGISTIGNMDGYP  
GRGFFDGLTMSAGGCTVTTVAENIGWAMTKVSTLIFVIAAMALIGMSPK  
FTALINFGALGATSTLAALVLYQLGIGRSDS"  
/complement(3675..4331)  
/gene="bme17"  
/complement(3675..4331)  
/gene="bme17"  
/codon\_start=1  
/transl\_table=11  
/product="Bme17"  
/protein\_id="AAL27668.1"  
/db\_xref="GI:16740534"  
/translation="MIDLCALANPORLELLCGLGSHDDIEICGGTLLHLETGTL  
HVWCVLNGEQRRAEAEARDLFGMDRFLQHVASFADSYFPAESRAIKEWLIAQR  
GRQPDIVFTHGDAHQDRHTVNLTNLFRDLQILEYIIPKMDGLGRPNAYSALP  
ASIMERKIDLLAKHFGTQRKDFGTFETFRGLARLRGMECRAPETVAEFAHARKLRF  
"  
/complement(4328..5137)  
/gene="bme20"  
/complement(4328..5137)  
/gene="bme20"  
/codon\_start=1  
/transl\_table=11  
/product="Bme20"  
/protein\_id="AAL27669.1"  
/db\_xref="GI:16740535"  
/translation="MKVILFCGLGTRIREYSENIPKPMIPLGHQPIILHHVMEYSDY  
GHNDFVLCIGYKANVVDFFLTIRPQTADFCDVSDGGRNVQLLEBVDWDSVTLIDTG  
IWRNIGIERLWSARSHVNNEDIIFLANSGLSDVDLDNMTQRFASGKACFLAVRPPL  
TVHLADISGQVRAPTNTSDIINGGVFLFKEIPDYMEEGELVLEPESRLAE  
NQLMAKYDGFWRSDTLDMQTLDEMVEKGMFPMKHKAEQRAANIAL"  
/complement(5140..5441)  
/gene="bme21"  
/complement(5140..5441)  
/gene="bme21"  
/note="putative C-methyl transferase"  
/codon\_start=1  
/transl\_table=11  
/product="Bme21"  
/protein\_id="AAL27670.1"  
/db\_xref="GI:16740536"  
/translation="MTIQADISRDHSTVQARTGCRGLKLRHSFVLDGMSPPCESF  
VBAETRLHEDNPPYPLHAQVCHFLVQLKEFSPADICTEYAFVPSFATSWYSHAKY  
CBEITRLHEDNPPYPLHAQVCHFLVQLKEFSPADICTEYAFVPSFATSWYSHAKY  
VDFPFGSLALEMIGEGKADLIGNNVLAQVPLDNDVFRGMQLLLKPGVITLPPHL  
VNLIQNDQDTIYEHFSPSLITIRFQAHHKLVKIDVNEELFTGGSRVTLAHGGS  
KRKQGRVSLAKREESFGEISITTEYQAEKTRRTKRDLLSFLIAAKNAGKRICGY  
APKGNTLLNYCGIGDFLGTVDNPNYPKHGFTPMHPIYDVSAINDNRPDYLLIL  
PWNFKDEIRIQWQHVVGEAKFIIPHPVTLIDPALVTEER"  
/complement(5731..7285)  
/gene="bme22"  
/complement(5731..7285)  
/gene="bme22"  
/note="putative epimerase"  
/codon\_start=1  
/transl\_table=11  
/product="Bme22"  
/protein\_id="AAL27671.1"  
/db\_xref="GI:16740537"  
/translation="MQKCNRGQAHFSAHQHRRVHVRAGTAKDERGWFARVYCEAA

FAERKLETHFPQHSLSPSRBKTLRGLHYQNEPHSETKLVTCQGIWVDVLDLRPHS  
PTVRWAGVLSAENGVLQYIPBCAGHGFOTLDDVLIRYMSKPYADPAAAGIRYDD  
PALDIPWPKSPVISQDLGWPL"  
complement(7248..8561)  
/gene="bme23"  
/complement(7248..8561)  
/gene="bme23"  
/codon\_start=1  
/transl\_table=11  
/product="Bme23"  
/protein\_id="AAL27672.1"  
/db\_xref="GI:16740538"  
/translation="MAAPQDPAIHLARLVPICRSLAGDVGRETLAIGEYLPPLH  
IHEVPTGTFLYDKNAPOEWLIRACIADLAGRIVDPARNLHVPMVSRMSEMSLD  
ALKNHITLFDQDLIPIRYTCYHGSWGFCAHNAALLMEDGFEDVVIDAERDGSIV  
FGFIRHGQTDATFLLSAHCPHSLANDNCSGALLALLGALKEAMKRRITRYLLRPGP  
"  
Query Match 15.6%; Score 78; DB 1; Length 36502;  
Best Local Similarity 51.1%; Pred. No. 0.11;  
Matches 183; Conservative 0; Mismatches 175; Indels 0; Gaps 0;  
Qy 140 TCACGTGTACGCGGGGACGTACAGGATCCACTCTCCGCCAGCCCGCGGAACCTCGCT 199  
Db 5172 TCAAGGTGACATGGGGAATGGGAATGATGAATTTCCGCGCCCATTCACGACATGCTGCA 5231  
Qy 200 CTTTCGCCCATGATCTCGTCGCGGTGTTCCAGGCGAAGACAGCGCGTAGTCCACCGCGT 259  
Db 5232 TCTCGCGGATAATCTCGTCCTTGAATTCATGCGCAGATGAGAATGTAGTCGGGCCGT 5291  
Qy 260 CGGCGGTGAACGGTCCGGGTGCGCACCGGATGTCGTCGCGGGGTGAGCCGCGCT 319  
Db 5292 AATTGTGATCGTCGCGAACAATCGTAAATCGGAATGTGATTCACAGGCGTAAAGCGCGCT 5351  
Qy 320 GTTTGGCGGGTGTCTGTCGCACACCCAGGAGACAGGTCCGGACCGATCGCGCAGAAGT 379  
Db 5352 GCTTTAGGATTCGGATCCACGGTGAACCCAGAAATCCGTGCGCATGCCCAATAGT 5411  
Qy 380 TCGTCACGGTGCGCTCTTTCGCGTCCGCGGTACGCCACACCGCTTGCCTTCGCGCT 439  
Db 5412 TCAAAAGCGTGTTCCTCCCTTTCGCGGGCGCCATAGCCACAGATGCGTTTCCCGCATTT 5471  
Qy 440 TGACCGAGTTTCAGCAGGCGCAGCAGGTGCGTGGGATGCCCTCGACCTCGCGCGCGAA 497  
Db 5472 TGGCGGCATCAGGAAGACAGAAATACAGTTCGTCGCGCGCTTTTCGCGCAA 5529  
RESULT 48  
AF329398/c 42291 bp DNA linear BCT 13-DEC-2002  
LOCUS Streptomyces roseochromogenes subsp. oscitans clorobiocin  
DEFINITION biosynthetic gene cluster, complete sequence.  
ACCESSION AF329398  
VERSION AF329398.1 GI:24940583  
KEYWORDS  
SOURCE  
ORGANISM Streptomyces roseochromogenes subsp. oscitans  
Streptomyces roseochromogenes subsp. oscitans  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE  
AUTHORS Pojer,F., Li,S.M. and Heide,L.  
TITLE Molecular cloning and sequence analysis of the clorobiocin  
biosynthetic gene cluster; new insights into the biosynthesis of  
aminocoumarin antibiotics  
Microbiology 148 (Pt 12), 3901-3911 (2002)  
JOURNAL 12480894  
PUBMED 2 (bases 1 to 42291)  
AUTHORS Pojer,F., Li,S. and Heide,L.  
TITLE Direct Submission  
JOURNAL Submitted (14-DEC-2000) Pharmazeutische Biologie, Pharmazeutisches  
Institut, Auf der Morgenstelle 8, Tuebingen 72076, Germany  
FEATURES  
source  
1..42291  
/organism="Streptomyces roseochromogenes subsp. oscitans"



/transl\_table=11  
/product="clog"  
/protein\_id="AA65221.1"  
/db\_xref="GI:24940594"  
/translation="MHTAVIIGTGMIGTSIGLSKQGVDTYVLMDSVVALRISBAVG  
AGVAEPPEAVLAVVAVPAAQVAVIASHOSRQARFYVDVAVGVKVSRLREBALGC  
DIATVVGHPPLVGRPGSPPLAADDLPDRPWALVPAATDASALNCALRELVLCGAI  
PVLDABARALAGTLVPOVLTLLAARLTADSIARLLGLGWNSRIPOLIGVDSI  
TSWTQALANAPVVSLEKLSLDLITLLETLRAVADSGSPAEPPDGHLLRFIRGID  
GNNRIPGRYGTETGLADIDVFVEDRPELARLPDRVASAGVVMRGINISQRPDSDLD  
RVTTSVTPRKAEILLHLRERKWPADA"  
/gene="clog"  
11746..12705  
11746..12705  
/gene="clog"  
/codon\_start=1  
/transl\_table=11  
/product="regulatory protein clog"  
/protein\_id="AA65222.1"  
/db\_xref="GI:24940595"  
/translation="MSNSGDEBITISPKITGRKRVISIGLLPPSKLLRSGESTEHI  
RALATEETLPPIVHREPRVIDGMHRLRAARFPGDELIEVVFVDSGSAADVFLAVE  
LNRAGLEPTMAERSAAARINDSPHNSDRKIARTTGLASTIASLRSSISGTVR  
RTGQDRIRPDGTGQRASDLAQNPASREVAKIARISVGTASDVRARLRGEP  
AVTARQAVKLKRPAAATQRTSPDYRVLSLRKQPSLRFTDVGRRLLDLGAPGS  
VFQIAQIADSVPEHCRIVVVDWARECAAAWQHLADQLAGRDTA"  
/gene="clog"  
12920..13135  
12920..13135

Query Match 15.6%; Score 78; DB 1; Length 42291;  
Best Local Similarity 51.1%; Pred. No. 0.1;  
Matches 183; Conservative 0; Mismatches 175; Indels 0; Gaps 0;  
QY 140 TCACGTGTACCGGGGACGTACAGATCCACTGTCCGCGCAGCCCGCGGAACCTCTGCT 199  
DB 37841 TCACGGTGGTTCGGGTATGGGACATATAACGGCGCGCGTTTCCAGAACGACTTCT 37782  
QY 200 CTTTGGCCGATGCTCGTCGGCGTGTTCAGGGGAGACAGGGCGGTATGTCACCGCGT 259  
DB 37781 CTTTGGTGTATCTCCGGCGGTAGTTCAGGCCAACAGAGTAGTAGTCCGGTACAT 37722  
QY 260 CGGGCGTGAACGGTCCGGGTGCGCACCCGAGATGCTGCGCGGGGTAGCGCGGCCCT 319  
DB 37721 GGTGCGTGCCTTTCGGTGACCATCGAAATCTGGACCTGGTAGATTTTCCCT 37662  
QY 320 GCTTGGCCGCGTGTCTGCGACACCCAGGACCGAGTCCGACCGATGCCGCGAAGT 379  
DB 37661 GTTTACAGAGTGTGTGTCTACATACAGAACCGGATCTCTGATGCCCGCGCATCGG 37602  
QY 380 TCGTCACGGTGGCGCTCTTCGCGTCCGCGGTAGCCACACCGCTTGCCTTCGCGCT 439  
DB 37601 CGAGCATGCTACTCCCTTTAGTTGGTCCGCGGTATCCAGCACATGATTTGCCCTTCCCA 37542  
QY 440 TGACCGAGTTTCAGCAGGCGCAGGTCGGTCCGGATGCCCTCGACCTCGCGCGCGAA 497  
DB 37541 CCACGGCGCATACCACTTCCTCGGATCTGTGCGGAACCTCGTTCGTACGCTCGCGGAA 37484

RESULT 49  
AX707118/c 42291 bp DNA linear PAT 04-APR-2003  
LOCUS Sequence 4 from Patent WO03014352.  
DEFINITION AX707118  
ACCESSION AX707118  
VERSION AX707118.1 GI:29563423  
KEYWORDS Streptomyces roseochromogenes  
SOURCE Streptomyces roseochromogenes  
ORGANISM Streptomyces roseochromogenes  
REFERENCE 1  
AUTHORS Heide, L. and Li, S.M.  
TITLE Nucleic acids for aminocoumarin biosynthesis

JOURNAL Patent: WO 03014352-A 4 20-FEB-2003;  
Universitaet Tuebingen (DE)  
FEATURES Location/Qualifiers  
source 1. 42291  
/organism="Streptomyces roseochromogenes"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:67357"  
ORIGIN  
Query Match 15.6%; Score 78; DB 6; Length 42291;  
Best Local Similarity 51.1%; Pred. No. 0.1;  
Matches 183; Conservative 0; Mismatches 175; Indels 0; Gaps 0;  
QY 140 TCACGTGTACCGGGGACGTACAGATCCACTGTCCGCGCAGCCCGCGGAACCTCTGCT 199  
DB 37841 TCACGGTGGTTCGGGTATGGGACATATAACGGCGCGCGTTTCCAGAACGACTTCT 37782  
QY 200 CTTTGGCCGATGCTCGTCGGCGTGTTCAGGGGAGACAGGGCGGTATGTCACCGCGT 259  
DB 37781 CTTTGGTGTATCTCCGGCGGTAGTTCAGGCCAACAGAGTAGTAGTCCGGTACAT 37722  
QY 260 CGGGCGTGAACGGTCCGGGTGCGCACCCGAGTCCGACCGATGCCGCGAAGT 319  
DB 37721 GGTGCGTGCCTTTCGGTGACCATCGAAATCTGGACCTGGTAGATTTTCCCT 37662  
QY 320 GCTTGGCCGCGTGTCTGCGACACCCAGGACCGAGTCCGACCGATGCCGCGAAGT 379  
DB 37661 GTTTACAGAGTGTGTGTCTACATACAGAACCGGATCTCTGATGCCCGCGCATCGG 37602  
QY 380 TCGTCACGGTGGCGCTCTTCGCGTCCGCGGTAGCCACACCGCTTGCCTTCGCGCT 439  
DB 37601 CGAGCATGCTACTCCCTTTAGTTGGTCCGCGGTATCCAGCACATGATTTGCCCTTCCCA 37542  
QY 440 TGACCGAGTTTCAGCAGGCGCAGGTCGGTCCGGATGCCCTCGACCTCGCGCGCGAA 497  
DB 37541 CCACGGCGCATACCACTTCCTCGGATCTGTGCGGAACCTCGTTCGTACGCTCGCGGAA 37484

RESULT 50  
SCO939126 295150 bp DNA linear BCT 11-FEB-2003  
LOCUS Streptomyces coelicolor A3(2) complete genome; segment 23/29.  
DEFINITION AL039126 AL009199 AL023517 AL023702 AL034415 AL034446 AL034492  
ACCESSION AL035212 AL035478 AL049727 AL132824 AL132997 AL589708 AL591522  
AL645882  
VERSION AL939126.1 GI:24413894  
KEYWORDS Streptomyces coelicolor A3(2)  
SOURCE Streptomyces coelicolor A3(2)  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Bentley, S.D., Chater, K.F., Cerdano-Tarraga, A.M., Challis, G.L.,  
Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,  
Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,  
Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,  
Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,  
Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S.,  
Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,  
Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrall, B.G.,  
Parkhill, J. and Hopwood, D.A.  
TITLE Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)  
JOURNAL Nature 417 (6885), 141-147 (2002)  
MEDLINE 21996410  
PubMed 12000953  
REFERENCE 2 (bases 1 to 295150)  
AUTHORS Bentley, S.D.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces  
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
COMMENT On or before Oct 26, 2002 this sequence version replaced

```

gi:20520744, gi:20520662, gi:20520663, gi:20520667, gi:20520669,
gi:20520761, gi:20520816, gi:20520870, gi:20520820, gi:20520846,
gi:20520823, gi:20520863, gi:20520869.
FEATURES             Location/Qualifiers
     1..295150
     /organism="Streptomyces coelicolor A3(2)"
     /mol_type="genomic DNA"
     /strain="A3(2)"
     /db_xref="taxon:100226"
     512..515
     /note="possible RBS upstream of SC7B7.02"
RBS
     524..1933
     /gene="SC06005"
     /note="synonym: SC7B7.02"
     524..1933
     /gene="SC06005"
     /note="SC7B7.02, possible lipoprotein, len: 469; has
similarity to signal sequence of Mycobacterium leprae
hypothetical protein TR:E334742 (EMBL:298604) MLCB2052.27
(445 aa), fasta scores; opt: 75 z-score: 199.2 E():
0.00062, 23.3% identity in 335 aa overlap. Contains
N-terminal signal sequence and appropriately positioned
PS00013 Prokaryotic membrane lipoprotein lipid attachment
site"
     /codon_start=1
     /transl_table=11
     /product="putative lipoprotein"
     /protein_id="CAA15782.1"
     /db_xref="GI:2661678"
     /db_xref="GOA:O50499"
     /db_xref="SPTREMBL:O50499"
     /translation="MTIRAGSLDRITLLRGAITATAMGSFAVACSPSSDKSDSGP
KGEKSNPFAGANSTVEAAIFDGYGTDYDVYANQVLGKQVQKVPVVDIAP
QLOPRFGNGPDLIDNSGDOIIGLILQLELDDLFEASTVEGKIAIDIVPGVK
DPGTFKDKVFNALNVYGVYKTLFEENGWTPPKTWDEALDLGQAAKGGKYLFW
HGKAARYRLLIDSAIKGGDEVRLEALEKGCWSPHVGQVVKWETVYKQKMF
VPGSGGTFQKAQAIWNDQKALLPSCGHENEMKATKADQMTGIPGSMITLDEPA
LPEYALRAAGPEFPIVQKKNPAGKEVRLMLSEKAAANFSKTLPIATVIGTVPA
DGYGSTALVSTQMLEAAGTINIFMYFVETVGLNTDQLVPWNSFLAGDLDDKGLTSL
QKISDKVREDDSDVKVYS"
misc_feature
     581..613
     /gene="SC06005"
     /note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
     1941..3002
     /gene="SC06006"
     /note="synonym: SC7B7.03"
     1941..3002
     /gene="SC06006"
     /note="SC7B7.03, probable integral membrane transport
protein, len: 353 aa, similar to many members of the MALFG
subfamily of the binding-protein-dependent transport
system membrane proteins eg. MSMP_STRMU Q00750 multiple
sugar-binding transport system permease protein (290 aa),
fasta scores; opt: 393 z-score: 636.8 E(): 2.6e-28, 32.5%
identity in 280 aa overlap. Contains PS00402
Binding-protein-dependent transport systems inner membrane
comp signature"
     /codon_start=1
     /transl_table=11
     /product="probable transmembrane transport protein"
     /protein_id="CAA15783.1"
     /db_xref="GI:2661678"
     /db_xref="GOA:O50500"
     /db_xref="SPTREMBL:O50500"
     /translation="MKDILPIATASRPPEAAGGRPRRKLIFDRVTFPLAFLGVP
LAIFVFLIPGQAFIWMGTDWRGFSDFNFGDFNFKMFDQDIFLAKURNVALLA
AFVPLVTLTALGVAITLGGPSKGPVRIGASFYRIISFFVYVPAIVGLIWAQ
MYDPNAGLNGVLTGLDQDFTFAWLGKAAAMPANVFIWGLVGFYAFVITAAIK
GVPELYEAKIDGAGRTTISITLPAIRDSVQTAYILGIALDAPFVYQAVVPGK
GPDNSTLTISQRLFNVAFAKQFGVATMGVLAATLVFAALFVLNRLTGCGEGES
KKAQSGRRARRAAKGAR"
misc_feature
     2589..2675

```

```

/gene="SC06006"
/note="PS00402 Binding-protein-dependent transport systems
inner membrane component signature"
2988..2992
/gene="SC06006"
/note="possible RBS upstream of SC7B7.04"
2999..3907
/gene="SC06007"
/note="synonym: SC7B7.04"
2999..3907
/gene="SC06007"
/note="SC06007"
/note="SC7B7.04, probable integral membrane transport
protein, len: 301 aa, similar to many members of the MALFG
subfamily of the binding-protein-dependent transport
systems membrane proteins eg. TR:Q06111 (EMBL:X66092)
putative transport system permease protein from
Clostridium perfringens (275 aa), fasta scores; opt: 590
z-score: 796.5 E(): 0, 33.6% identity in 265 aa overlap.
Contains PS00402 Binding-protein-dependent transport
systems inner membrane comp signature"
     /codon_start=1
     /transl_table=11
     /product="probable transmembrane transport protein"
     /protein_id="CAA15784.1"
     /db_xref="GI:2661679"
     /db_xref="GOA:O50501"
     /db_xref="SPTREMBL:O50501"
     /translation="MSVIAADKPKASDRKPVSDRKLAFASASDRRFAAISHALLILW
SVIIVPMLVMSKFSKTBILSPFSLPDHMFENYANAWTANIGKFLNSVIVV
VSALTLMLGAMCAVLAIRFEFGRRLLIYVMLAGLTFPVPLAIVPFOQLNFGLL
NTRFGLIITYAPALPTFMFLYSFERSLPHDVVEALIDGAGDWRAFPQVMLPMWRP
GNAVAIFNYFGLMNGDFLNPALNDQKWLVTQGMAYASSQYVDIDYGALFAAIVV
TVPVLLVYFCVQRRRIGSVSQGTR"
3554..3640
/gene="SC06007"
/note="PS00402 Binding-protein-dependent transport systems
inner membrane comp signature"
3935..3984
/note="hairpin loop with 23bp stem"
4128..4131
/note="possible RBS upstream of SC7B7.05"
4142..5353
/gene="SC06008"
/note="synonym: SC7B7.05"
4142..5353
/gene="SC06008"
/note="SC7B7.05, probable transcriptional repressor
protein, len 403 aa; similar to many eg. XYLX_ARATH Q44406
xylose repressor from Anaerocellum thermophilum (399 aa),
fasta scores; opt: 548 z-score: 767.3 E(): 0, 33.1%
identity in 393 aa overlap. Contains PS01125 ROK family
signature and helix-turn-helix motif at aa positions 29-50
(score 2016, +6.05 SD)"
     /codon_start=1
     /transl_table=11
     /product="probable transcriptional repressor protein"
     /protein_id="CAA15785.1"
     /db_xref="GI:2661680"
     /db_xref="GOA:O50502"
     /db_xref="SPTREMBL:O50502"
     /translation="METFGQSLSLHRLNLRVVRVRLAGSLTQAEIARTTGLSAATV
SNIVRELKGGTVETPTPSAGGRKARSVLSGDAGIVIGVDGHTFLRVAVGNLAHOV
LABESEPLDQVSSQGFDAQOLVRLIEATGVDRAKIAVGGVPGPIDVSGTILG
STAALFGWGIIRPABELRGRLGVPVHNDNALGALGELVWGGRGVRLDIYIKVASG
VGAALVINGKIYRPGGTAGEIHTLDEAGFVRCGRCGLTETFAARYVLPLQPS
HGTLDMEGVVRLLARDPDGCRRLVADVGRHIGSGVNLCLNLSPRVVLGGDLAEAG
EALVGPISRVGCRYAIPSAARQLSVLPGALGGAELGALALSEMGSSTLLDGSAT
GALPAATPAFT"
4784..4867
/gene="SC06008"
/note="PS01125 ROK family signature"
5462..5466
/note="possible RBS upstream of SC7B7.06"

```

```
gene      5475. .6590
/ gene="SCO6009"
/ note="synonym: SC7B7.06"
CDS       5475. .6590
/ gene="SCO6009"
/ note="SC7B7.06, probable solute-binding protein of
transmembrane transport system, len: 403 aa; similar to
many eg. XYL ECOLI P37387 d-xylose-binding periplasmic
protein precursor (330 aa), fasta scores; Opt: 619
z-score: 573.4 E(): 8.9e-25; 35.7% identity in 325 aa
overlap. Contains N-terminal signal sequence and
appropriately positioned PS00013 Prokaryotic membrane
lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="solute-binding protein"
/protein_id="CAA15786.1"
/db_xref="GI:2861661"
/translation="NNTRMRAAVALTAMAVSLACGSAKESDGGSDSDSAAKKG
DDIKVGLLLPENQTAHYEKFKPMTEKKVKELTNKGEVVYANAKQDASLQNOQVDITM
Query Match      12.8%; Score 64; DB 1; Length 295150;
Best Local Similarity 51.5%; Pred. No. 3.8; Indels 6; Gaps 1;
Matches 176; Conservative 0; Mismatches 160;
QY 141 CACGTGTACGCGGGAGCTACAGGATCCACTGTCCGCCAGCCCGCGGAACTCCTGCTC 200
Db 187772 CAGGTGCAGCCGGTTCGCGCAGCGGTCAGGATCCGTGCTCGAGGACGCTCCAG 187831
QY 201 CTTGCCATGATCTCGTCGGCTGGTTCAGCGGAGAGCAGCGGTAGTCCACCGCTC 260
Db 187832 CGTCGGGATCAGTCGCGCGGTGGCGGCTGCTAGTCCCGACGGGTCCAGCAGCG 187891
QY 261 GGGCGTGAACGGCTTCGCGGTGCGCACCGGGATGTGCGTCCCGGGGTGAGCGGCCCTG 320
Db 187892 GCGGTGAAGCGCGCGGACGCTGTCGGGACGAA-----GGGAGCAGCAGCACGTG 187945
QY 321 CTTGCCGCGTGTGTCGACACCCAGGAGACGAGTCCGAGACCGATGCCGAGAGTT 380
Db 187946 CGAGGCCAGCTCCTGTGGCCGCGCCGCGCAGACCCCGCGGGGCGCGCGGCGACCCG 188005
QY 381 CGTCAGGTGGCGCTCTTCGCGTCCGCGCTACGCCACCCCGCTTGCCCTCGGCCTT 440
Db 188006 GCGCGGTGCGGGCTCTCTCAGCGCCCGCGCAGCCCTTCGCGAGTGCACGCGCGC 188065
QY 441 GAGCGAGTTCAGAGGCGAGCAGGTGCGTGCAGTGCCTC 482
Db 188066 GCTGACCGAGCGTCAACGCGCCGCTCGCCGCTCCAGGCCGC 188107
```

Search completed: June 27, 2004, 19:39:06  
Job time : 2371.22 secs



QY 121 ACACCGGGGTGACGTGCTCAGCTGTAACGGGGGAGCGTACAGGATCCACTGTCGCCCA 180  
Db 30366 ACACCGGGGTGACGTGCTCAGCTGTAACGGGGGAGCGTACAGGATCCACTGTCGCCCA 30307  
QY 181 GCCCGGGGAACTCTGCTGCTTCCGCAATGATCTGCTCGGGGTGTTTCCAGGCGAAGAGC 240  
Db 30306 GCCCGGGGAACTCTGCTGCTTCCGCAATGATCTGCTCGGGGTGTTTCCAGGCGAAGAGC 30247  
QY 241 AGCCGCTAGTCCACCGGCTGGGGTGAAACGCTGCGGGGTGCGCACCGGGATGTGCGTG 300  
Db 30246 AGCCGCTAGTCCACCGGCTGGGGTGAAACGCTGCGGGGTGCGCACCGGGATGTGCGTG 30187  
QY 301 CGGGGGGTAGCCGGCCCTGCTTCGGCGGCTGCTGTCGACACCCAGGAGACCAAGTCC 360  
Db 30186 CGGGGGGTAGCCGGCCCTGCTTCGGCGGCTGCTGTCGACACCCAGGAGACCAAGTCC 30127  
QY 361 GGACCGATGCGCGCAAGATTGCTCACGGTGGCGCTCTTCGCGCTGCGCCGCTAGCCACC 420  
Db 30126 GGACCGATGCGCGCAAGATTGCTCACGGTGGCGCTCTTCGCGCTGCGCCGCTAGCCACC 30067  
QY 421 ACCCGCTGCTCGGCTGCTGAGTTCAGCGAGTTCAGCGGGGAGCAGTCCGTCGGATGCC 480  
Db 30066 ACCCGCTGCTCGGCTGCTGAGTTCAGCGAGTTCAGCGGGGAGCAGTCCGTCGGATGCC 30007  
QY 481 TCGAGCTCGGGCGGAACCT 500  
Db 30006 TCGAGCTCGGGCGGAACCT 29987  
RESULT 2  
US-09-758-759-1  
; Sequence 1, Application US/09758759  
; Publication No. US20040101832A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tim X.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Evernimicin Biosynthetic Genes  
; FILE REFERENCE: ID0983K US  
; CURRENT APPLICATION NUMBER: US/09/758,759  
; CURRENT FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/175,751  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea  
US-09-758-759-1  
Query Match 100.0%; Score 500; DB 12; Length 109519;  
Best Local Similarity 100.0%; Pred. No. 2e-113;  
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGTACCCGACCGGTGTCGCCGAAACAGAGTCCAGATACCGCGGAGAGAACACCCCGGT 60  
Db 1 GGTACCCGACCGGTGTCGCCGAAACAGAGTCCAGATACCGCGGAGAGAACACCCCGGT 60  
QY 61 AGTCCGGGTAGACGGTGGGCGGAAAGCGTACCGGCTTCGACCGGTTCGACCGGTGGGGGG 120  
Db 61 AGTCCGGGTAGACGGTGGGCGGAAAGCGTACCGGCTTCGACCGGTTCGACCGGTGGGGGG 120  
QY 121 ACACCGCGGGTTCAGTCTGTCAGTGTACCGGGGAGCTACAGGATCCACTGTCGCCCA 180  
Db 121 ACACCGCGGGTTCAGTCTGTCAGTGTACCGGGGAGCTACAGGATCCACTGTCGCCCA 180  
QY 181 GCCCGGGGAACTCTGCTGCTTCCGCAATGATCTGCTCGGGGTGTTTCCAGGCGAAGAGC 240  
Db 181 GCCCGGGGAACTCTGCTGCTTCCGCAATGATCTGCTCGGGGTGTTTCCAGGCGAAGAGC 240  
QY 241 AGCCGCTAGTCCACCGGCTGGGGTGAAACGCTGCGGGGTGCGCACCGGGATGTGCGTG 300

Db 241 AGCCGCTAGTCCACCGGCTGGGGTGAAACGCTGCGGGGTGCGCACCGGGATGTGCGTG 300  
QY 301 CGGGGGGTAGCCGGCCCTGCTTCGGCGGCTGCTGTCGACACCCAGGAGACCAAGTCC 360  
Db 301 CGGGGGGTAGCCGGCCCTGCTTCGGCGGCTGCTGTCGACACCCAGGAGACCAAGTCC 360  
QY 361 GGACCGATGCGCGCAAGATTGCTCACGGTGGCGCTCTTCGCGCTGCGCCGCTAGCCACC 420  
Db 361 GGACCGATGCGCGCAAGATTGCTCACGGTGGCGCTCTTCGCGCTGCGCCGCTAGCCACC 420  
QY 421 ACCCGCTGCTCGGCTGCTGAGTTCAGCGAGTTCAGCGGGGAGCAGTCCGTCGGATGCC 480  
Db 421 ACCCGCTGCTCGGCTGCTGAGTTCAGCGAGTTCAGCGGGGAGCAGTCCGTCGGATGCC 480  
QY 481 TCGAGCTCGGGCGGAACCT 500  
Db 481 TCGAGCTCGGGCGGAACCT 500  
RESULT 3  
US-10-107-431-232/c  
; Sequence 232, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTH  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 232  
; LENGTH: 1251  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea africana  
US-10-107-431-232  
Query Match 73.8%; Score 369; DB 13; Length 1251;  
Best Local Similarity 100.0%; Pred. No. 4.1e-81;  
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 132 TCAGCTGCTCAGCTGACCGGGGAGCTACAGATCCACTGTCGCCGACCCCGCGGAA 191  
Db 1251 TCAGCTGCTCAGCTGACCGGGGAGCTACAGATCCACTGTCGCCGACCCCGCGGAA 1192  
QY 192 CTCCTGCTCTTCGCCATGATCTCGTGGCGTGGTTCCAGGCGAAGAGACGCGTAGTC 251  
Db 1191 CTCCTGCTCTTCGCCATGATCTCGTGGCGTGGTTCCAGGCGAAGAGACGCGTAGTC 1132  
QY 252 CACCGCTGGGGGTGAAACCGCTCGGGGTGCGCACCGGATGTCGTCGCGGGGTGAG 311  
Db 1131 CACCGCTGGGGGTGAAACCGCTCGGGGTGCGCACCGGATGTCGTCGCGGGGTGAG 1072  
QY 312 CCGGCCCTGCTTCGGCGGCTGCTGTCGACACCCAGGAGACCAAGTTCGGAACCGATGCC 371  
Db 1071 CCGGCCCTGCTTCGGCGGCTGCTGTCGACACCCAGGAGACCAAGTTCGGAACCGATGCC 1012  
QY 372 GCAGAGTTCGTACCGTGGCGCTCTTCGCGCTGCGGCGCTACGCCACACCCGCTTGGC 431  
Db 1011 GCAGAGTTCGTACCGTGGCGCTCTTCGCGCTGCGGCGCTACGCCACACCCGCTTGGC 952  
QY 432 CTCGGCTTCAGCGAGTTCAGCGGGGAGCAGTCCGTCGGATGCCCTTCGACGTCGCG 491  
Db 951 CTCGGCTTCAGCGAGTTCAGCGGGGAGCAGTCCGTCGGATGCCCTTCGACGTCGCG 892  
QY 492 GCGGAACCT 500  
Db 891 GCGGAACCT 883

## RESULT 4

US-09-758-759-2/c  
; Sequence 2, Application US/09758759  
; Publication No. US20040101832A1  
; GENERAL INFORMATION:  
; APPLICANT: Hosted, Thomas J.  
; APPLICANT: Wang, Tim X.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Evernimycin Biosynthetic Genes  
; FILE REFERENCE: ID0983K US  
; CURRENT APPLICATION NUMBER: US/09/758,759  
; CURRENT FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/175,751  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1248  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1248)  
; OTHER INFORMATION: evda  
US-09-758-759-2

Query Match 73.2%; Score 366; DB 12; Length 1248;

Best Local Similarity 100.0%; Pred. No. 2.2e-80; Mismatches 0; Indels 0; Gaps 0;  
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	135	GCTGTCACGTGTACGCGGGGACGTACAGGATCCACTGTTCGCCAGCCCGCGGAATC	194
DB	1248	GCTGTCACGTGTACGCGGGGACGTACAGGATCCACTGTTCGCCAGCCCGCGGAATC	1189
QY	195	CTGCTCTTCGCATGATCTCTGCGGCTGTTCCAGGCGAAGACAGCGCGTAGTCCAC	254
DB	1188	CTGCTCTTCGCATGATCTCTGCGGCTGTTCCAGGCGAAGACAGCGCGTAGTCCAC	1129
QY	255	CGCGTCGGGGGTGAACCGCTCCGCGGTGCGCACCGGAGTGTGCTGCCGGGGGTGAGCG	314
DB	1128	CGCGTCGGGGGTGAACCGCTCCGCGGTGCGCACCGGAGTGTGCTGCCGGGGGTGAGCG	1069
QY	315	GCCTGCTTGGCGGGGTCTGTGCGACACCGAGGACAGAGTCGGGACCGATGCCGA	374
DB	1068	GCCTGCTTGGCGGGGTCTGTGCGACACCGAGGACAGAGTCGGGACCGATGCCGA	1009
QY	375	GAAGTTCGTACAGGTGGCGCTCTTCGCCGTCGCGCGGTACGCCACCGCTTGCCTC	434
DB	1008	GAAGTTCGTACAGGTGGCGCTCTTCGCCGTCGCGCGGTACGCCACCGCTTGCCTC	949
QY	435	GGCCTTGAGCGAGTTACGAGGGGAGCAGGTGCGGATGCCCTCGACGTCGCGGC	494
DB	948	GGCCTTGAGCGAGTTACGAGGGGAGCAGGTGCGGATGCCCTCGACGTCGCGGC	889
QY	495	GAACCT 500	
DB	888	GAACCT 883	

## RESULT 5

US-10-107-431-282/c  
; Sequence 282, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; TITLE OF INVENTION: BIOSYNTHETIC LOCI  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28

; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 282  
; LENGTH: 10035  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea aurantiaca  
US-10-107-431-282

Query Match 59.7%; Score 298.4; DB 13; Length 10035;  
Best Local Similarity 76.1%; Pred. No. 5e-64;  
Matches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;

QY	1	GGTACCGACCGTGTCCCGGAACACGAGTCAGATACGCGGAGAGAACACCCCGGT	60
DB	3742	GATATCCGAGCGTCTCGGTGAAGACCGAGTCCAGTACGGGACACAGAACCCCGCG	3683
QY	61	AGTCCGGGTACACGCTGGGCGCGAAGCGCTACCGCCTTCAGCGTCAGCGGGCGGCG	120
DB	3682	GGTCCGGGAAGACCCCGGGGTGAAGACGTACGGCGCTCGACCGCGAGTTCGGTTGTA	3623
QY	121	ACACC---GGGGGTGAGCTCGTACGTGTACGGGGGAGCGTACAGGATCCACTGCC	176
DB	3622	CCTGACGCGCCCACTCAATCCGACGTGAACGTGCGGAACGTACAGGATCCAGGCC	3563
QY	177	GCCAGCCCGCGGAACTCTCTCTTCGCCATGATCTCGCGCTGGTTCAGAGGAA	236
DB	3562	CCCGCCTGCGGAAAGCGCTCTCTTCGCCATGATCTCGCGCTGGTTCAGAGGAA	3503
QY	237	GAGCAGCGGTACTCCACCGCTCGGGGTGAACCGTCCGGGTGCGCACCGGATGTG	296
DB	3502	GAGCAGCGGTACTCCACCGCTCGGGGTGAACCGTCCGGGTGCGCACCGGATGTG	3443
QY	297	CGTCCCGGGGTGAGCGCGCCCTCTTGGCCGGCGCTCGTTCGCACACACCCAGAG	356
DB	3442	CGTCCCGGGGTGAGCGCGCCCTCTTGGCCGGCGCTCGTTCGCACACACCCAGAG	3383
QY	357	GTCGCGACCGATGCCGCGAAGTTCGTCACGTCGCGCTCTTCGCGTCGCGCGTACGC	416
DB	3382	GTCGCGACCGATGCCGCGAAGTTCGTCACGTCGCGCTCTTCGCGTCGCGCGTACGC	3323
QY	417	CACACCGCTTTCCTTCGCGCTTCGAGCGAGTTCAGCAGGCGGAGCAGGTGCGTCCG	476
DB	3322	CACACCGCTTTCCTTCGCGCTTCGAGCGAGTTCAGCAGGCGGAGCAGGTGCGTCCG	3263
QY	477	GCCCTCGACGTGCGCGGGAAC 498	
DB	3262	CCGCGCGACCTGGCGGAGAAC 3241	

## RESULT 6

US-09-769-734-49/c  
; Sequence 49, Application US/09769734  
; Publication No. US2003014366A1  
; GENERAL INFORMATION:  
; APPLICANT: Ecopia Biosciences Inc.  
; TITLE OF INVENTION: Genetic Locus for Evernimycin Biosynthesis  
; FILE REFERENCE: PA 005-US  
; CURRENT APPLICATION NUMBER: US/09/769,734  
; CURRENT FILING DATE: 2001-01-26  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 49  
; LENGTH: 11115  
; TYPE: DNA  
; ORGANISM: M. carbonacea  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (8)..(1207)  
; OTHER INFORMATION: ORF 41 (positive strandedness)  
; OTHER INFORMATION: incomplete: C-terminus only  
; NAME/KEY: misc.feature  
; LOCATION: (1213)..(2331)  
; OTHER INFORMATION: ORF 42 (positive strandedness)



1 GGTACCCGACCGTGTCCCGGAACAACGAGTCGAGATACGCGCGAGAGGGAACACCCCCCGGT 60

Db 12022 GGTACCCGACCGTGTCCCGGAAACAACGAGTCGAGATACGGCGAGAGGAACAACCCCGCGGT 12081  
QY 61 AGTCCGGGTAGACGGTGGCGCGGCGAAGCGGTACCGCGCTTCGACGGTCAAGCGGCGGCGG 120  
Db 12082 AGTCCGGGTAGACGGTGGCGCGGCGAAGCGGTACCGCGCTTCGACGGTCAAGCGGCGGCGG 12141  
QY 121 ACACCGCGGG 131  
Db 12142 ACACCGCGGG 12152

## RESULT 9

US-10-107-431-220/c  
; Sequence 220, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; TITLE OF INVENTION: BIOSYNTHETIC LOCI  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 220  
; LENGTH: 603  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea africana  
US-10-107-431-220

Query Match 24.8%; Score 124; DB 13; Length 603;  
Best Local Similarity 100.0%; Pred. No. 3e-21;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTACCCGACCGTGTCCCGGAAACAACGAGTCGAGATACGGCGAGAGGAACAACCCCGCGGT 60  
Db 124 GGTACCCGACCGTGTCCCGGAAACAACGAGTCGAGATACGGCGAGAGGAACAACCCCGCGGT 65  
QY 61 AGTCCGGGTAGACGGTGGCGCGGCGAAGCGGTACCGCGCTTCGACGGTCAAGCGGCGGCGG 120  
Db 64 AGTCCGGGTAGACGGTGGCGCGGCGAAGCGGTACCGCGCTTCGACGGTCAAGCGGCGGCGG 5  
QY 121 ACAC 124  
Db 4 ACAC 1

## RESULT 10

US-10-107-431-278/c  
; Sequence 278, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; TITLE OF INVENTION: BIOSYNTHETIC LOCI  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 278  
; LENGTH: 15240  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea africana  
US-10-107-431-278

Query Match 23.6%; Score 118.2; DB 13; Length 15240;  
Best Local Similarity 53.3%; Pred. No. 4.6e-20;  
Matches 249; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 33 AGATACCGCGAGAGGAACAACCCCGGTAGTTCGGGTAGACGGTGGCGGGAAGAGCGGTAC 92  
Db 12684 AGAGATGATCCAGCGGAGCAGCAGCATGTGTCCGGGACCGGGCTTCGGGGTCACTCATGA 12625  
QY 93 GCGCTTTCGACGGTCAAGCGGCGGCGGACACCGCGGGTCACTCGTCACTGTCAGGTAGCGG 152  
Db 12624 AAGTTGATCAGTCTCGGTGGGGGGGTGTATCCCGGTCCGGCGACCGTCTCGACCCG 12565  
QY 153 GGGACGTPACAGGATCCATGTCCGCGAGCCCGCGGAACTCTCTGTCTCTTCGCGCATGAT 212  
Db 12564 GGGCAGCGGAAAGATCAACCGGTGCGGTGCGGCGAGCATCTCCGCGCTCCCGGGGACGAT 12505  
QY 213 CTGCTCGCGGTGTTCCAGGCGAAGAGCAGCGGTAGTCCACCGCGTCCGGCGTGAACGC 272  
Db 12504 CTCGTCCCGAAGTGCACAGGCGAGCAGCATAGTAGTATCCGGGCGGGCGCGCGCATC 12445  
QY 273 GTCGGGGTGCACACCGGATGTGCTGCCGGGGGTGAGCGGCCCTTGTGGCGCGGT 332  
Db 12444 CTGCTCGCTGATGATTTTCGATGTCCGTGCGGAGGTGCGCGGCCACCTTGTCCGGATT 12385  
QY 333 GGTGTCGACACACCGAGGAGCAGGTCCGGACCGATGCGCAGAACTGCTCAGGTGGC 392  
Db 12384 GCGCTCCCGCGGTACCGGATGAGTCCGGTCTATGCCGCGAGAACTGACAGAGGTGTT 12325  
QY 393 GCTCTTCGCGTCCGCGCGGTACGCCACACCGCGCTTGCCTTCGCGCTTTCAGCGAGTTCA 452  
Db 12324 GCCCTTCGTGCGAGCGCGGTAGACGTGACCGTGGCGGCTTGGCGCGCGAGCTCGCGC 12265  
QY 453 CAGGCGGAGAGTCCGTGCGGATGCCCTTCGACGTCCGGCGGCGGAC 499  
Db 12264 CAGGCGGCTCACCTGTCACCGTGTGGCGCACCTGTCGGCGAAGC 12218

## RESULT 11

US-09-758-759-50/c  
; Sequence 50, Application US/09758759  
; Publication No. US20040101832A1  
; GENERAL INFORMATION:  
; APPLICANT: Hosted, Thomas J.  
; APPLICANT: Wang, Tim X.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Evernimycin Biosynthetic Genes  
; FILE REFERENCE: ID0983K US  
; CURRENT APPLICATION NUMBER: US/09/758,759  
; CURRENT FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/175,751  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 50  
; LENGTH: 1236  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1236)  
; OTHER INFORMATION: evim  
US-09-758-759-50

Query Match 22.6%; Score 112.8; DB 12; Length 1236;  
Best Local Similarity 56.5%; Pred. No. 1.5e-18;  
Matches 210; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 128 CGGTTCAGTCTCGTACCGGGGAGCGTACAGGATCCACTGTCCGCCAGCCCGC 187  
Db 1234 CGGTTCAGTCTCGTACCGGGGAGCGTACAGGATCCACTGTCCGCCAGCCCGC 1175  
QY 188 GGAACCTCTCTCTTCGCGCATGATCTCGTCCGGTGTTCAGGCGAAGAGCAGCGGT 247  
Db 1174 GCATCTCCCGCTCCCGGGGAGCATCTCGTCCCGGAAAGTCCAGGCGAGACCATGAT 1115  
QY 248 AGTCCACCGGTCCGGCGGTGAACGCGTCCGGGTGGCACCGGGATGTCCGTCCCGCGG 307



```
; TYPE: DNA
; ORGANISM: Micromonospora carbonacea aurantiaca
US-10-107-431-120

Query Match:      18.5%; Score 92.4; DB 13; Length 1290;
Best Local Similarity 53.6%; Pred. No. 1.4e-13;
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 142 ACCTGTACGCGGGGAGCTACAGGATCCACTGTCCGCCAGCCCGGGGAACTCTCTCTCC 201
DB 1250 ACCCGCAGGCTGGGACGGGAAGATCAGCTTGTTCGGCCGCCACCGTGGCGCTCG 1191
QY 202 TTGCGCATGATCTCGTCGGCGTGTTCAGGCGAAGACGAGCCGCTAGTCCACCGGCTCG 261
DB 1190 CGCGCCACGATCTCGTCGTGGAAGTCCACGCGAGCACACGAGTAGTGTTCGGCGCGCG 1131
QY 262 GGGGTGAACCGCTCCGGGGTGGCGCACCGGATGTGGCTGCCGGGGGTGAGCCGGCCCTCG 321
DB 1130 GCCCGCAGTCCGCTCGCTGATGATCTCGATGTCGTACCGAGGGTCCGGCGCGGACC 1071
QY 322 TTGCGCGCGCTGCTGTGTCGACACCCAGGAGACAGGTCGCGACCGATGCCGCAAGATTTC 381
DB 1070 TTGTCCGGGTTCGGCTCGCGGGCGGTACGGGATCAGCGTGGCGTGCATCCCGCAGTACTGC 1011
QY 382 GTACAGGTGCGCTCTTCGCGCTCGCGCGCTGACCCACACCGCTTGGCCCTCGGCTTG 441
DB 1010 AGAGGGGTGTCCCTTGTGGAGGCGCCGTAGACGTGCACGGTCTTCGCTGTGCGC 951
QY 442 AGCGAGTTCAGCGGGCGAGCGAGTGGTCCGATGCCATGCCCTCGACGTGGCGGCGAACC 499
DB 950 AGACCATGCAGCATCTTACGAGTTCTCGCGGTTCGCGGTGCGCCCGCACGTTGTGCGGGAACC 893

RESULT 15
US-09-769-734-8/c
; Sequence 8, Application US/09769734
; Publication No. US2003014366A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia Biosciences Inc.
; TITLE OF INVENTION: Genetic Locus for Everninomicin Biosynthesis
; FILE REFERENCE: PA 005-US
; CURRENT APPLICATION NUMBER: US/09/769,734
; CURRENT FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 14252
; TYPE: DNA
; ORGANISM: M. carbonacea
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (459)..(1280)
; OTHER INFORMATION: ORF 5 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (2677)..(3747)
; OTHER INFORMATION: ORF 7 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (1280)..(2566)
; OTHER INFORMATION: ORF 6 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (3899)..(4774)
; OTHER INFORMATION: ORF 8 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (4893)..(5303)
; OTHER INFORMATION: ORF 9 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (5365)..(6306)
; OTHER INFORMATION: ORF 10 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (6350)..(7204)
; OTHER INFORMATION: ORF 11 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (7371)..(8198)
```

```
; OTHER INFORMATION: ORF 12 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (8304)..(9098)
; OTHER INFORMATION: ORF 13 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (9462)..(10493)
; OTHER INFORMATION: ORF 14 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (10665)..(11384)
; OTHER INFORMATION: ORF 15 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (11387)..(12700)
; OTHER INFORMATION: ORF 16 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (12971)..(14185)
; OTHER INFORMATION: ORF 17 (negative strandedness)
US-09-769-734-8

Query Match:      18.5%; Score 92.4; DB 10; Length 14252;
Best Local Similarity 53.6%; Pred. No. 9.3e-14;
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 142 AGTGTACGCGGGGACGTACAGGATCCACTGTCCGCCAGCCCGGGGAACTCTCTCTCC 201
DB 2529 ACCCGCAGGCTGGGACGGGAAGATCAGCTTGTTCGGCGCGCACCGTGGCGCGCTCG 2470
QY 202 TTGCGCATGATCTCGTCGGCGTGTTCAGGCGAAGACGAGCGGTAGTCCACCGCTCG 261
DB 2469 CGCGCCACGATCTCGTCGTGGAAGTCCACCGGAGCACCATGATGTGTCGGCGCGCG 2410
QY 262 GCGGTGAACCGCTCCGGGGTGGCGCACCGGATGTGCTGCCGGGGGTGAGCGGCGCTTC 321
DB 2409 GCCCGCAGTGGCGCTCGCTGATGATCTCGATGTCGTACCGAGGGTCCGGCGCGGACC 2350
QY 322 TTGCGCGCGCTCGTGTGCAACCCAGGAGACAGGTCCGCGACCATGCGCAGAGTTTC 381
DB 2349 TTGTCCGGGTTCGCTCGCGCGCGTACCGGATGACGCTGACAGGTCTTCGCTGCGCGC 2290
QY 382 GTCACGCTGGCGCTCTTCGCGCTCCGCCGTACGCCACACACCGCTTCCCTCTCGGCTTC 441
DB 2289 AGGAGGGTGTTCCTTGTGGAGGCGCGGTAGACGTGCACGGTCTTCGCTGCGCGC 2230
QY 442 AGCGAGTTCAGCGGGCGAGCGAGTGGTGGGATGCGCTCGAGTCCGCGCGGCGAACC 499
DB 2229 AGACCATGCAGCATCTTACGAGTTCTGTCGGGTGCGCCCGCACGCTTGTGCGGGAACC 2172

RESULT 16
US-10-107-431-280/c
; Sequence 280, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Stafifa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORF
; FILE REFERENCE: 3001-705
; CURRENT APPLICATION NUMBER: US/10/107,431
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 280
; LENGTH: 48221
; TYPE: DNA
; ORGANISM: Micromonospora carbonacea aurantiaca
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: "n" at position 106 is undetermined and represents a or g or t c
; OTHER INFORMATION: c
US-10-107-431-280

Query Match:      18.5%; Score 92.4; DB 13; Length 48221;
```

Best Local Similarity 53.6%; Pred. No. 7.6e-14; Mismatches 192; Conservative 0; Gaps 0; Indels 0;

QY 142 AGGTGACGCGGGGAGCTACAGGATCCACTGCGCGAGCCGCGGAACTCTGCTCC 201  
Db 8244 ACCCGAGGCTGGGACGGGAGATCAGTTGGTTCCGGCGCCACCGTGGCGCTCG 8185

QY 202 TTGCGCATGATCTCGTGGCGTGTTCAGGCGAAGAGAGCGGTAGTCCACCGCTCG 261  
Db 8184 CGCGCCACGATCTCGTGGGAAGTCCACGCGCACGAGTAGTGTCCGGCGCGG 8125

QY 262 GCGGTGAACGCTCGCGGTGCGCACCGGATGTCGTCGCGGGGTGAGCGGCCCTCG 321  
Db 8124 GCCCGGAGTCCGCTCGTATGATCTCGATGTCGACGAGGTCCGGCGCGGAC 8065

QY 322 TTGCGCGGCTCGTGTGCGACACCGAGGAGACAGTCCGACCGATCGCGAGATTC 381  
Db 8064 TTGTCGGGTTGCGCTCGCGCGGTACGCGATCAGCGTCCGTCGATCCCGAGTATCG 8005

QY 382 GTACGGTGGCTCTTCCCGTCCGCGGTAGGACACCGCTTCGCGCTCGCGCTTG 441  
Db 8004 AGAGGGTGTCCCTTGTGGAGCGCGGTAGAGCTGACGCTCTCCGCTGTCGCG 7945

QY 442 AGCGAGTTCAGCAGGCGGAGCGTGGTGGGATGCGCTCGACGTCGCGCGGAC 499  
Db 7944 AGACCATGAGCATCTTGACGATTCTGTCGCGTGGCGCGACGTTGTCGCGAACC 7887

RESULT 17  
US-10-107-431-277/c  
; Sequence 277, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; FILE OF INVENTION: BIOSYNTHETIC LOCI  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 277  
; LENGTH: 45055  
; TYPE: DNA  
; ORGANISM: Streptomyces mobaraensis  
US-10-107-431-277

Query Match 18.1%; Score 90.6; DB 13; Length 45055;  
Best Local Similarity 52.5%; Pred. No. 2.1e-13; Mismatches 198; Conservative 0; Gaps 0; Indels 0;

QY 122 CACCGGGGTACGCTCGTACGCTGACCGGGGAGTACAGGATCCACTGTCGCGCAG 181  
Db 9974 CCGCGACCGGTGACACCACTTCCAGCGGGGAGCGGAGACAGGTTGCGCGCG 9915

QY 182 CCGCGGGAATCTCGTCTGCTTCCGATGATCTGTCGCGGTGTTCCAGGCGAAGCA 241  
Db 9914 CCGCGGGAATCTCGTCTGCGGCGACCATCGCTCCGCGTAGATCCACGCGGAGCA 9855

QY 242 GCGGTAGTCCACCGGTGGGGTGAACGCGTCCGGGTGCGCACCGGATGTGCTGC 301  
Db 9854 GGAACGGTCCGGTCCAGCGCCCGGGCTCTCTTCGAGAGCATCGGATGTCCGTGC 9795

QY 302 CGGGGTGAGCGCGCTGCTTCCGCGGGTGTGTCGACACCCAGGAGACAGTCCG 361  
Db 9794 CGGGGTGAGCAGCGCAATCTGCTTCACTTCCGCGGTGACAGCGGAGCGTCT 9735

QY 362 GACCGTCCGCGAGATCTGTCACGTCGCGCTCTTCCGCTCGCGCTAGCCACCA 421  
Db 9734 CGTCCAGCGCGATCTGAGGATGACGTTGCTCCCTTGTGGAGGCGCGTACCCGAG 9675

QY 422 CCGCTTGGCCCTCGCGCTTACGAGCTTCAGAGGGGAGAGGTCGCTGCGGATGCCCT 481  
Db 9674 TGGCGACGCCCTTGTCCCGCGAGGTGTGAGGAAGTCCACGAGCCGGTCCCGGATTCCT 9615

QY 482 CGACGTCGCGCGGGAAC 498  
Db 9614 CGTCCCGCGCGGAAC 9598

RESULT 18  
US-10-107-431-118/c  
; Sequence 118, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; FILE OF INVENTION: BIOSYNTHETIC LOCI  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 118  
; LENGTH: 1224  
; TYPE: DNA  
; ORGANISM: Streptomyces mobaraensis  
US-10-107-431-118

Query Match 17.8%; Score 87.8; DB 13; Length 1224;  
Best Local Similarity 52.6%; Pred. No. 1.8e-12; Mismatches 191; Conservative 0; Gaps 0; Indels 0;

QY 136 CTGTCACGCTGACGCGGGGAGCTACAGGATCCACTGTCGCGCAGCCCGCGGAACTCC 195  
Db 1223 CACACCACTTCCAGGGCGGACAGGTTGCGCGCGCGCGCCAGGAGTCCG 1164

QY 196 TGTCTCTTCCGCTGATCTCGTGGCGTGTTCAGGCGAAGAGAGCGGTAGTCCAC 255  
Db 1163 CGTTCCCGGCGACCATCGCTCCCGTAGATCCACGGGAGGACCAAGAACCGGTCCGCG 1104

QY 256 GCGTCGCGGTGAACGCTCGCGGTGCGCACCGGATGTCGCGCGGGGTGAGCGG 315  
Db 1103 TCCAGCGCCCGGCTCTCTTCGAGACGATCGGATGTCGCGCGGTGAGGAG 1044

QY 316 CCCTGCTTGGCGCGCTGCTGTCGACACCCAGGAGACAGGTCGCGACCGATGCGCGCAG 375  
Db 1043 CCGAACCTTCTCTTCACTCGCGATGACAGGAGGCTCTCTCTCCAGGCGCGAG 984

QY 376 AGTTCTGCTACGCTGGGCTTTCGCGCTGCGCGCTAGCCACCGCTTTCCTCG 435  
Db 983 TACTGGAGGATGAGTTCGCTTGGTGGAGCGCGCTTACCCGAGGTGCGAGCCCTTG 924

QY 436 GCCTTGAGCGATTCAGCAGGCGGAGCAGGTTCGCTGCGGATGCTTCGAGTCCGCGGCG 495  
Db 923 TCCCGCGAGTGTGAGGAAGTCCACGAGCGGTTCGCGGATTCCTTCGCTCCGCGCGG 864

QY 496 AAC 498  
Db 863 AAC 861

RESULT 19  
US-10-229-148B-1/c  
; Sequence 1, Application US/10229148B  
; Publication No. US20040091975A1  
; GENERAL INFORMATION:  
; APPLICANT: Meiji Seika Kaisha, Ltd.  
; TITLE OF INVENTION: Midcamycin biosynthetic genes  
; FILE REFERENCE: 138451 US  
; CURRENT APPLICATION NUMBER: US/10/229,148B  
; CURRENT FILING DATE: 2002-08-28

```

; PRIOR APPLICATION NUMBER: 210516/2002
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 1
; LENGTH: 84428
; TYPE: DNA
; ORGANISM: Streptomyces mycarofaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((11)..(675))
; OTHER INFORMATION: ORF42 (fragment)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((1168)..(2202))
; OTHER INFORMATION: ORF41
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((2220)..(3215))
; OTHER INFORMATION: ORF40
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((3237)..(4691))
; OTHER INFORMATION: ORF39
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((4695)..(5948))
; OTHER INFORMATION: ORF38
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6048)..(6629))
; OTHER INFORMATION: ORF37
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6653)..(7945))
; OTHER INFORMATION: ORF36
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8149)..(9015)
; OTHER INFORMATION: ORF35
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9012)..(9335)
; OTHER INFORMATION: ORF34
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9328)..(10458)
; OTHER INFORMATION: ORF33
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10521)..(11603)
; OTHER INFORMATION: ORF32
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11729)..(12961)
; OTHER INFORMATION: ORF31
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((13016)..(14044))
; OTHER INFORMATION: ORF30
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((14074)..(15096))
; OTHER INFORMATION: ORF29
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15643)..(17466)
; OTHER INFORMATION: ORF28
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((17522)..(18895))
; OTHER INFORMATION: ORF27
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19063)..(20229)
; OTHER INFORMATION: ORF26
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((20307)..(21743))
; OTHER INFORMATION: ORF25
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((21733)..(22527))
; OTHER INFORMATION: ORF24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((22534)..(23571))
; OTHER INFORMATION: ORF23
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((23555)..(24463))
; OTHER INFORMATION: ORF22
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((24460)..(25650))
; OTHER INFORMATION: ORF21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((25647)..(26105))
; OTHER INFORMATION: ORF20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((26180)..(27391))
; OTHER INFORMATION: ORF19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27937)..(28983)
; OTHER INFORMATION: ORF18
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29244)..(42779)
; OTHER INFORMATION: ORF1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42823)..(48657)
; OTHER INFORMATION: ORF2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (48712)..(59802)
; OTHER INFORMATION: ORF3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59850)..(64556)
; OTHER INFORMATION: ORF4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64687)..(70365)
; OTHER INFORMATION: ORF5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70365)..(71078)
; OTHER INFORMATION: ORF6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71113)..(72360)
; OTHER INFORMATION: ORF7
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72400)..(73665)
; OTHER INFORMATION: ORF8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73694)..(75043)
; OTHER INFORMATION: ORF9
; FEATURE:
; NAME/KEY: CDS
```

LOCATION: Complement((75899)..(76570))  
 OTHER INFORMATION: ORF10  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: Complement((76602)..(77765))  
 OTHER INFORMATION: ORF11  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (78039)..(79313)  
 OTHER INFORMATION: ORF12  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: Complement((79391)..(81052))  
 OTHER INFORMATION: ORF13  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (81541)..(82356)  
 OTHER INFORMATION: ORF14  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (82760)..(83362)  
 OTHER INFORMATION: ORF15  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: Complement((83495)..(84142))  
 OTHER INFORMATION: ORF16  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (84329)..(84428)  
 OTHER INFORMATION: ORF17 (fragment)  
 US-10-229-148B-1

Query Match 16.8%; Score 84.2; DB 17; Length 84428;  
 Best Local Similarity 51.2%; Pred. No. 7e-12;  
 Matches 197; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

QY 114 CGGGCGGACACGGGGGTGACGTGTCACGTGTACGGGGGACGTACAGATCCACTG 173  
 DB 12982 CGGGCGGGCGGGCGGCGCTGTGCACACGACGTGACGTGCGGCGACGGGAAGAGAGGTTT 12923

QY 174 TCCGCGAGCCCGCGGAACTCTGCTCTTCCCATGATCTGCTCGGGGTGTTCCAGGC 233  
 DB 12922 GCGCGCGGCTCCCGGTAGGCCGCTCCCGCTCAGGGAAGCGTCCGATAGATCCACGG 12863

QY 234 GAAGAGACGCGGTAGTCCACCGCTCGGGCGGTGAACCGCTCGGGGGTCCGACCGGGAT 293  
 DB 12862 CAGCACCAGCAGTTGGTTCGGGCTTGAGGAGCTTGCCCTCTCTCGGACACGATCGGGAT 12803

QY 294 GTGCGTGGCGGGGTGAGCGGCCCTGCTTGCGCGGCGTGTGCGCACACCCAGGAGAC 353  
 DB 12802 CGCGATCCGGGGGTGAAGCGGCCGCTTCTCTGCTGACCTCGCCGATGCGGGCAG 12743

QY 354 CAGGTCGACCGATGCGCGAGAGTTGCTACCGGTGGCGCTCTTCCGCGTCCGCGCGTA 413  
 DB 12742 GTCCGCTCGGTGAGACCGCAGTACTGGAGGATCAGTTGCCCTTGGTGGAGGCGCCGTA 12683

QY 414 CGCACACACCGCTTCCCTCGGCTTGAGCGAGTTACGAGGGGAGCAGTGGTGGTCCG 473  
 DB 12682 TCACCGGTGAGACGCGCCGCTCGGGGAGCGGTGAGGAAGTGCATCAGGGGCTCCG 12623

QY 474 GATGCGCTCAGCTCGGGCGGAAC 498  
 DB 12622 CTGGTCTGACACGCGCGCGAAC 12598

RESULT 20  
 US-10-084-846A-1/c  
 ; Sequence 1, Application US/10084846A  
 ; Publication No. US2004006026A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WEITNAUER, GABRIELE  
 ; APPLICANT: MUHLENWEG, AGNES  
 ; APPLICANT: TREFZER, AXEL

APPLICANT: BECHTHOLD, ANDREAS  
 TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
 FILE REFERENCE: 1974-005  
 CURRENT APPLICATION NUMBER: US/10/084,846A  
 CURRENT FILING DATE: 2003-02-25  
 PRIOR APPLICATION NUMBER: PCT/EP01/09815  
 PRIOR FILING DATE: 2001-08-24  
 PRIOR APPLICATION NUMBER: DE 101 09 166.4  
 PRIOR FILING DATE: 2001-02-25  
 NUMBER OF SEQ ID NOS: 120  
 SOFTWARE: PatentIn Ver. 3.2  
 SEQ ID NO 1  
 LENGTH: 59816  
 TYPE: DNA  
 ORGANISM: Streptomyces viridochromogenes  
 US-10-084-846A-1

Query Match 16.2%; Score 81; DB 16; Length 59816;  
 Best Local Similarity 50.9%; Pred. No. 4.5e-11;  
 Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGGTACGCTGTCAGTGTACCGGGGACGTACAGGATCCACTGTCCGCCAG 181  
 DB 15020 CTCGCCACCCGATCACAACTTCACGCTGGGCAAGCGAAGACAGGCTGCTCCG 14961

QY 182 CCGCGCGAACTCTGCTCTTCCCATGATCTGCTCGCGTGTTCAGGCGAAGACA 241  
 DB 14960 AGGCCAGGAAGTCGCTTCCCGGCGACCATCGGTCCTCCATCCACGCGAGACCA 14901

QY 242 GCGGTAGTCCACCGCTCGGGGTGAACGCTCGGGGTGCGACCGGGATGCGTGC 301  
 DB 14900 GGAACGTGTCGGGCTCAAGCGCCCGGCTCTCTTCGAGACGATCGGGATGTCGTGC 14841

QY 302 CGGGGTGAGCGCGCCCTGCTTGGCGGGTGTGTCGCGACACCCAGGAGACAGGTCG 361  
 DB 14840 CCGCGGTGAGCAGCGCAACTTGTCTCTTCACTCGCGATGACGCGAGGCTCT 14781

QY 362 GACCGATGCGGAGAGTTCGTACCGTGGCTCTTCCCGTCCGCGGTAGCCACCA 421  
 DB 14780 CGTCCAGCGCGAGTACTGCGATGACGTGCGCTTCGTCGAGGCGCGTAGCCAGGG 14721

QY 422 CCGCTTGCCTCGGCTTTCAGCGAGTTTCAGAGGCGGAGAGGTCGTCGCGATCCCT 481  
 DB 14720 TGTGACGCCCTTGTCCCGGCGGCTGAGGATTCAGGATCGTCCAGGACTCT 14661

QY 482 CGACGTGCGGCGGAAC 498  
 DB 14660 CGGTCCCGCGCGAAC 14644

RESULT 21  
 US-10-084-846A-2  
 ; Sequence 2, Application US/10084846A  
 ; Publication No. US2004006026A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WEITNAUER, GABRIELE  
 ; APPLICANT: MUHLENWEG, AGNES  
 ; APPLICANT: TREFZER, AXEL  
 ; APPLICANT: BECHTHOLD, ANDREAS  
 TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
 FILE REFERENCE: 1974-005  
 CURRENT APPLICATION NUMBER: US/10/084,846A  
 CURRENT FILING DATE: 2003-02-25  
 PRIOR APPLICATION NUMBER: PCT/EP01/09815  
 PRIOR FILING DATE: 2001-08-24  
 PRIOR APPLICATION NUMBER: DE 101 09 166.4  
 NUMBER OF SEQ ID NOS: 120  
 SOFTWARE: PatentIn Ver. 3.2  
 SEQ ID NO 2  
 LENGTH: 59816  
 TYPE: DNA  
 ORGANISM: Streptomyces viridochromogenes

US-10-084-846A-2

Query Match 16.2%; Score 81; DB 16; Length 59816;  
Best Local Similarity 50.9%; Pred. No. 4.5e-11;  
Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGGTCACTGCTCAGCTGTACGCGGGGACGTACAGATCCACTGTCCGCCAG 181  
DB 44797 CTCGCCACCCGATCACACAATTCAGAGTGGGACGCGAAGACAGGCTGCTCCG 44856

QY 182 CCGCGGGAAGTCTGCTCTTCCGCAATGATCTGCTGGGCTGGTTCAGAGGAGCA 241  
DB 44857 AGCGCAGGAAGTCCGCTTCCGCGGCGACCATCCGCTCCCGGTAGATCCACGGGAGACCA 44916

QY 242 GCGGTAGTCCACCGCTCCGGGGGTGAACCGCTCCGGGGTGCACACCGGATGTGGGTGC 301  
DB 44917 GGAATGCTCGGGTCAAGCGCCCGGGCTTCTCTCCGAGAGATCGGGATGTCTGTC 44976

QY 302 CGGGGGTACCGCGCTTGTGGCGGCGTCTGTGCGACACCCAGGAGACAGGTCCG 361  
DB 44977 CGGGCGGTGACGACGCGCAACTTGTCTCTGTTCACTCGGCGATGCGAGGCGGTCT 45036

QY 362 GACCGATGCGCAGAGTCTGTCACGCTGCGCTTTCGCGTCGCGCGTACGCCACCA 421  
DB 45037 CBTCCAGCGCGAGTACTGAGATGACGTGTCCTTCTGCGAGGCGCGTACCGAAGG 45096

QY 422 CCGCTTGCCTCGGCTTGAAGAGTTCAGCAGGCGAGCAGGTGCGGTCCGATCCCT 481  
DB 45097 TGTCCAGCCCTTGTCCGCGAGCGGTGAGGAATCCAGGAGTCCGTCACGGGACTCT 45156

QY 482 CGAGCTCGCGGGGAAC 498  
DB 45157 CGGTCCCGCGGGCAAC 45173

## RESULT 22

US-10-084-846A-43/c  
; Sequence 43, Application US/10084846A  
; Publication No. US20040006026A1  
; GENERAL INFORMATION:  
; APPLICANT: WEITNAUER, GABRIELE  
; APPLICANT: MUHLNWEIG, AGNES  
; APPLICANT: TREFFZER, AXEL  
; APPLICANT: BECHTHOLD, ANDREAS  
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084,846A  
; CURRENT FILING DATE: 2003-02-25  
; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
; PRIOR FILING DATE: 2001-02-25  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patent in Ver. 3.2  
; SEQ ID NO 43  
; LENGTH: 1224  
; TYPE: DNA  
; ORGANISM: Streptomyces viridochromogenes  
; FEATURE:  
; OTHER INFORMATION: avig1 dna: partial sequence of coding strand 1; nucleotide 1  
; OTHER INFORMATION: corresponds to nucleotide 13,784 of coding strand 1.

US-10-084-846A-43

Query Match 16.2%; Score 80.8; DB 16; Length 1224;  
Best Local Similarity 53.1%; Pred. No. 9.5e-11;  
Matches 172; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 175 CCGCAGCCCGGGGAACTCTGCTCTCCGATGATCTGCTCGCGGGTGTTCAGGGG 234  
DB 1184 CTTCCGAGGCGGAAAGTCCGCTTCCCGGCGCACCATCCGCTCCCGGTAGATCCACGG 1125

QY 235 AAGAGCAGCGGTAGTCCACCGGTGCGGGGTGAACCGTCCCGGGTGCACCGGATG 294

DB 1124 AGGACCAGGAACCTGTCGGGCTCAAGCGCCCGGCTCTCTCTCGAGACGATCGGGATG 1065

QY 295 TCGCTGCGGGGTGAGCCGCCCTGTTGCGCGGCGTCTGTCGACACCCAGGAGACC 354

DB 1064 TTCGTGCGGGCGGTGTAGCAGCCGAACCTGTCTCTGTTCACTCGGCGATCGAGGCGAGG 1005

QY 355 AGGTCCCGACCGATGCCGACGAAGTTGCTCACGGTGGCGCTCTTCGCGCTCGCGCCGTAC 414

DB 1004 AGCGTCTCGTCCAGGCCGCACTGTCAGGATGACGTTGCCCTTCGTCGAGCGCCGTAG 945

QY 415 GCCACACCCGCTGCTCGGCTTTCAGCGAGTTCAGCAGGCGAGCAGGTTCGGTTCGGG 474

DB 944 CCAAGGCGTGTGACGCCCTTGTCCCGGAGCGGTGAGGAATTCAGGAGTTCGGTTCACGG 885

QY 475 ATGCCCTCGACGTCCGGCGGAAC 498

DB 884 GACTCTCTCGGTCCGCGGGCGGAAC 861

## RESULT 23

US-10-107-431-218/c  
; Sequence 218, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING OR  
; TITLE OF INVENTION: BIOSYNTHETIC LOCI  
; FILE REFERENCE: 3001-705  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 218  
; LENGTH: 624  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea aurantiaca  
; US-10-107-431-218

Query Match 11.9%; Score 59.4; DB 13; Length 624;  
Best Local Similarity 71.6%; Pred. No. 1.8e-05;  
Matches 78; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 GGTACCCGACGCTGCCCGGAACACGAGTCAGATCGGCGAGGAGAACACCCCGGT 60

DB 127 GATATCCGAGCGTCTCGGTGAAGACCGAGTCAGGTACGGGACACGAGACCCCGG 68

QY 61 AGTCGGGTAGACGTCGGCGCGCGAGGCGTACGCGCTTCGACGTCAG 109

DB 67 GGTGCGGAAGACCGCGGGGTGAAGAGCTACGCGCCCTCGACGCGGAG 19

## RESULT 24

US-10-156-761-2058/c  
; Sequence 2058, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02



```
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 2058
/ LENGTH: 1671
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(1671)
US-10-156-761-2058

Query Match      11.4%; Score 57; DB 15; Length 1671;
Best Local Similarity 50.1%; Pred. No. 5.8e-05;
Matches 172; Conservative 0; Mismatches 165; Indels 6; Gaps 1;

QY 141 CACGRTACGGGGGAGCGTACAGGATCCACTGTCCGCGAGCCCGCGGAACTCCTGTC 200
DB 1563 CAGGTGGAGAGCGCGGGCACAGCGCTCCACGAAACCGTCCAGTCCAGAGAACGCTCCAG 2523112
QY 201 CTTGCCATGATCTCGTGGCGTGTTCACGCGGAAGAGCGGCGGTAGTCCACCGCTC 260
DB 2523111 GGTCCGGATCAGCTCGGCGCGTGGCGGCTGTCTCCCGCAGGGGTTCAGCAGAG 2523052
QY 261 GGGCGGTGAACGCGTCCGGGGTCCGACACCGGGATGTGCGTCCGGGGGTGAGCCGCCCTG 320
DB 2523051 GGGCGGTGAAGCGCGCGCACGCTCGTCCGGGACGAA-----GGGGAGCAGCAGGACGTG 2522998
QY 321 CTTGGCCGCGTGTGTGCGCACACAGAGAGACGAGTCCGACCCGATGCGCGCAGAGTT 380
DB 2522997 CAGAGGCGAGCTCTCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2522938
QY 381 CGTACGCGTGGCGCTCTTCGCGCGTCCGCGCGTACGCCACACCCCGCTTTCCTCGGCCCT 440
DB 2522937 GCGGGCGTCCGCGCGCTCTTCAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2522878
QY 441 GAGCGAGTTACAGAGGGCGAGCAGTCCGTGCGGATGCGCTCG 483
DB 2522877 GCTGACCCCGAGCGTGAGCGCGCGCGCTGCTTCAGGGCGCGCG 2522835

RESULT 26
US-10-156-761-4482/c
/ Sequence 4482, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRO
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ CURRENT FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 4482
/ LENGTH: 2982
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(2982)
US-10-156-761-4482

Query Match      11.2%; Score 56.2; DB 15; Length 2982;
Best Local Similarity 47.5%; Pred. No. 8.3e-05;
Matches 233; Conservative 0; Mismatches 253; Indels 5; Gaps 2;

QY 7 CGACCGGTGTCGCGGAAACAGTTCAGATACGCGGAGGAGAACACCCCGGTAGTCCG 66
DB 1087 CGGCGCGTCTCTTGAACCTCAGGTACGGGATGAGCGTGAACCGCATGTCTGACCG 1028
QY 67 GGTAGAGGTGGGCGGAGCGGTACGCGCTTCGACGTCACGGGCGCG--CGCGACAC 124
DB 1027 GGAAGTCCAGTGTGGGCGCGGTAGCGGCCCGCGCTGACTGGCGGAAAGCGGTCT 968
QY 125 CGCGGGTTCAGCTCGTCACTGTACGCGGGGAGCGTACAGGATCCATGTCGCCAGCCC 184
DB 967 CGGCGGCGAGTGGCGGCGACAGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 908
QY 185 GCGGGAACCTCTGCTCTCTTCGCGCGGTGTTCCAGGGCGAGGAGCAGCG 244
```

Db 907 CTTCTGCTACTGTCGAGCAGATGTCGCTTCTCGCGGAGAAATCGACGA 548  
QY 245 CTTAGTCCACCGGTCCGGGTGAACCGCTCCGGGTGCGCACCGGGATGTGCGTCCCG 304  
Db 847 CCAGGTCAACTGTCGGGCCCCCAGCGCGCGAGCGACACCGGCTCGGGCGCACGG 788  
QY 305 GGGTACCGCGCCTGTTGCGCGGCTGCTGTCGCACACCCAGGAGACCAAGTCCGGAC 364  
Db 787 G---CAGGTCCGCGTGGAGCGGCTTGGCGATCAGCCAGACCACTTCTCAGCGGTGAGCG 731  
QY 365 CGATCCGCGAAGTTCGTACCGTGGCGCTCTTTCGCGCTGCGCGCGGTACGCCACCAACC 424  
Db 730 GGGTCCCTTGGAGCCCTCAGGCTCTCGTGGACCTCGATCGCGACCTGCTCACTCGG 671  
QY 425 GCTTGCCTCGCTTACGAGTTCAGACGGGAGGAGGAGTGGTGGCGATCCCTCGA 484  
Db 670 CGAGGACGGAGGGGACAGGTACTGTCGCGGACCCCGGTGGCGGTGGAGACCGCCTGGT 611  
QY 485 CGTCGCGCGCG 495  
Db 610 CGCGCGCGCG 600

RESULT 27  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 11.2%; Score 56.2; DB 15; Length 9025608;  
Best Local Similarity 47.5%; Pred. No. 2.2e-05;  
Matches 233; Conservative 0; Mismatches 253; Indels 5; Gaps 2;  
QY 7 CGACCGTTCGCGAACAACGAGTCGAGATACGCGGAGAGGAACACCCCGGTAGTCGG 66  
Db 5495898 CGGCGCGGTCTTGAACCTCAGGTACGGGATGAGCGTGAACCGCATGTCGATCTCGACCG 5495957  
QY 67 GGTAGACGGTGGGCGGCGAAGCGGTACGCGCTTCGACGGTTCAGCGGCGG--CGCGACAC 124  
Db 5495958 GGAAGTCACGTGTCGGCGGCGTAGCGGCCCGCCAGCTGAGTCGCGGAACGGGTCT 5496017  
QY 125 CGCGGCTGACCTCGTCACTGTATACGGGGGAGAGTACAGGATCCACTGTCCGCCAGCCC 184  
Db 5496018 CGGCGGCGCACTTGGCGGCGACCAAGGCGTGGTGTAGTGTCTGCTGCTCGCCGCTGA 5496077  
QY 185 GCGGAACTCTGCTCTTCCGCAATGATCTCGTCGGCGGTGTTCCAGGCGAAGACGACG 244

Db 5496078 CTTCTGCTACTGTCGAGCAGATGTCGCTTCTCGCGGAGAAATCGACGA 5496137  
QY 245 CTTAGTCCACCGGTCCGGGTGAACCGCTCCGGGTGCGCACCGGGATGTGCGTCCCG 304  
Db 5496138 CCAGGTCAACTGTCGGGCCCCCAGCGCGCGAGCGACACCGGCTCGGGCGCACCG 5496197  
QY 305 GGGTACCGCGCCTGTTGCGCGGCTGCTGTCGCACACCCAGGAGACCAAGTCCGGAC 364  
Db 5496198 G---CAGGTCCGCGTGGAGCGGCTTGGCGATCAGCCAGACCACTTCTCAGCGGTGAGCG 5496254  
QY 365 CGATCCGCGAAGTTCGTACCGTGGCGCTCTTTCGCGCTGCGCGCGGTACGCCACCAACC 424  
Db 5496255 GGGTCCCTTGGAGCCCTCCAGGCTCTGTCGACCTCGATCGCGACCTGCTCACTCG 5496314  
QY 425 GCTTGCCTCGCTTACGAGTTCAGACGGGAGGAGGAGTGGTGGCGATCCCTCGA 484  
Db 5496315 CGAGGACGGAGGGGAGGAGGTACTGTCGCGGACCCCGGTGGCGGTGGAGACCGCCTGGT 5496374  
QY 485 CGTCGCGCGCG 495  
Db 5496375 CGGCGCGCGCG 5496385

RESULT 28  
US-10-282-122A-15092/c  
; Sequence 15092, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malore, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15092  
; LENGTH: 837  
; TYPE: DNA  
; ORGANISM: Bordetella pertussis  
US-10-282-122A-15092

```

Query Match      11.0%; Score 54.8; DB 13; Length 837;
Best Local Similarity 46.2%; Pred. No. 0.00022;
Matches 182; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY      106 TCAGCGGGCGGGCGGACACCGCGGGTCAGCTCGTCAAGTGTACGCGGGGACGTACAGG 165
DB      539 TCGACTGGCTGGCTGGCCACGATGCCGTATTTCGGCGTGTCTTCGGCGGCACCTCCCTGT 480

QY      166 ATCCACTGTCCGCCACGCGCGGAACTCCTGTCTCTTCGCCATGATCTCTCGCGCGTGG 225
DB      479 ACGCCCAAGCAGCTGTCCCTTGGCGCGCCACCGCCACGTCGATCACTGTCTTGAACACCGGC 420

QY      226 TTCAGAGCGAAGACAGCGCGTAGTTCACCCGCTCGGCGCTGAACGCTCCGGGTGGCG 285
DB      419 GTGTCCGCGTGCATCAGTGCATCGCCAGCAGCACCGCGAAGGCTCGTTGCCCTACCGCC 360

QY      286 ACGGGATGTGCTGCGCGGGGTGAGCGCGCCCTGCTTGGCGCGCTGCTGTGCGACACC 345
DB      359 GGCSCGGCGTCAACACGCGATGTCCAGGCCACGCGGGCGCATTTGCGGATGTACAGG 300

QY      346 CAGGAGACCAAGTCCGACCGATGCCGAGAAAGTTCTGTCAGGTGGCGCTCTTCGCCGTC 405
DB      299 CAGTTCACTGTGCGGGGAGGATGTGCGGGACCAAGCGCCAGCAGTTCGTGCTTGGCCCTTC 240

QY      406 GCGCGCTACGCCACACCGCTTGCCTTCGCGCTTGACGAGTTTCAGCAGGCGGAGCAGG 465
DB      239 GCTTCCAGTGGTTCAGTTCGGGCGCGATCGAAATGGTCTTCGATGGCCCGCTTG 180

QY      466 TCGGTGCGGATGCCCTCGACGTCCGCGCGCGAACC 499
DB      179 TTGGGCGCGGTGACGAAGATGAGTCCGTGATCC 146

RESULT 29
US-10-425-114-24834/c
; Sequence 24834, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 24834
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURES:
; OTHER INFORMATION: Clone ID: LIB3689-204-A2_FLI
US-10-425-114-24834

```

GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2271  
; LENGTH: 1137  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1137)  
US-10-156-761-2271

Query Match 10.7%; Score 53.6; DB 15; Length 1137;  
Best Local Similarity 49.3%; Pred. No. 0.00042;  
Matches 140; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 215 CGTCGGCGTGTCCAGCGAAGACGACGCGTGTGTCACCGCGTGGGGGTGAACGGCT 274  
Db 1030 CGTGGGGGAGCCACCGATGAATCCAGCGCGCGAGCACTCCAGGACCGGGCGCGA 971

QY 275 CCGGGGTGCGACCGGATGTGCGTCCGGGGGTGAGCGCGCCCTGCTTGGCGGGTGG 334  
Db 970 TCCGGCGACGGCTTCTGACAGTGTAGCGGGCGACCGAGGACCGGCGACCGACTCT 911

QY 335 TGTGCGACACCGAGGACAGGTCCGGACCGATGCCGAGAGTTTCGTACCGTGGCGC 394  
Db 910 CGTGGCGCGCTCCGTGCCGATCGCCCGTGCAGACACCTCGAGGAGCGCCAGGGCGCCT 851

QY 395 TCTTCCGCGTGGCGCGTACGCCACCGCTTGGCCCTCGCGCTTGGAGAGTTCAGCA 454  
Db 850 CCACTCGACCGGAGCTCGGCCGCTTCTGAGCGTGGCGCTCCCGCTCCAGCACT 791

QY 455 GCGGAGCAGGTGCGTGGGATGCCCTCGACGTCGCGCGGAAC 498  
Db 790 GCTCGAGAGCG 747

RESULT 32  
US-10-156-761-6854/c  
; Sequence 6854, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6854  
; TYPE: DNA

LENGTH: 540  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(540)  
US-10-156-761-6854

Query Match 10.7%; Score 53.4; DB 15; Length 540;  
Best Local Similarity 47.1%; Pred. No. 0.00053;  
Matches 235; Conservative 256; Indels 8; Gaps 2;

QY 5 CCGGACCGTGTCCCGAACAACAGTCCAGATACGCGCGCTTCGACCGTCCAGCGGGCGGGCGAC 124  
Db 510 CTTGGCCTCGTCCGCGAGCGTGGGCTCCCGCGCTCGGCTGGGTGGCATGGAGCGCGC 451

QY 65 CCGGTAGACGGTGGCGGGAAGCGTACGCGCTTCGACCGTCCAGCGGGCGGGCGGAC 124  
Db 450 CGCGGGACCGTCCGACCGTGGGATGCCGCTTGGAGCTCACCCTACGAGCGGCGT 391

QY 125 CCGCGGGTACGCTGCTACGCTGTCAGCGGG-----GGACGTACAGGATCCACTGTCGCC 179  
Db 390 CCGCTCGTCCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCG 331

QY 180 AGCCCGCGGAACCTCTCTCTCTCGCCATCATCTCTCGCTCGCGCTGTTCCAGGCGAAG 239  
Db 330 ACGCGGTGTCACCGCTGCTTGGTACTCCGAGCGCGGAGCCACCGCTCCGAGAA 271

QY 240 CAGCGGTAGTCCACCGCGTCCGCGTGAACCGTCCGCGGTGGCGTCCGCGGTGTCGCT 299  
Db 270 GCCAGGAGCGGTGGAAGTCCAGCGCGCTGACAGGCTTCGACACTGTGCGCGAG 211

QY 300 CCGGGGTGAGCGCGCGCTCTTGGCGGGTGTGTCGACACCGCGGAGCGGTCGCTAGACG 151  
Db 210 TCTTGGCGAGGCGGCGTCCGAGCGGAGCGGCTCCGTAGACGAGAACCGCTGGA 416

QY 360 CCGACCGATGCCGAGAGTTTCGTACCGTGGCGCTCTTCGCCGT---CGCGCGCTAGC 416  
Db 150 CCGCGCGAGCGGTGGCGGTAGAGCTTCCCCAACTGGCGGTGAGCGTGGCGCTGCG 91

QY 417 CACACCGCTTCCCTCGCGCTTGGCGGTGAGCGAGTTCCAGCGGGCGAGCGTGGTGGAT 476  
Db 90 GTCCACTCCCGCGGACCGCTCGATGTCCCGCACCGCAAGTGCAGGCTGGCGCGC 31

QY 477 GCCCTCGACGTCGGCGCG 495  
Db 30 CTGGCGCTCGTGGGTGG 12

RESULT 33  
US-10-156-761-6184/c  
; Sequence 6184, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6184  
; LENGTH: 894  
; TYPE: DNA

```
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(894)
US-10-156-761-6184

Query Match      10.7%; Score 53.4; DB 15; Length 894;
Best Local Similarity 48.8%; Pred. No. 0.00049;
Matches 144; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 191 ACTCTGCTCTCCCTCCGATGATCTGTCGGGGTGTTCAGCGAAGAGCAGCGTAGT 250
DB 508 ACCCGAGTTCTGGGACAGAGCGCGTCTCGCGAAGCGCTCGAGAAAGCAGCCAGCCG 449
QY 251 CCACCGCGTCCGGCGTGAACCGTCCGGGGTCCGACCGGATGTGCGTCCGGGGTGA 310
DB 448 CGCCAGATAGTCCCGCGGTGTCGGTGGAGATCTGTTGCGGGTGCCTGGGAGG 389
QY 311 GCGGCGCTGTTGCGCGGGTGTGTCGACACCCAGAGACAGGTCGAGACGATGC 370
DB 388 ACAGGTCCAGGCGGTGACAGGTACACGTGCATCACCCTGTCGCGCGCGCGCGATGT 329
QY 371 GCGAGAAGTTCGTCACGGTGGGGCTTTCGCGGTGCGCGCGTACGCCACACCGCTGTC 430
DB 328 CCGCGCGGTGGTGGGAGCGGTGCTTCAGTCTGATGAAGTACTCCGGCCCCCGGCGA 269
QY 431 COTCGGCTTGAGCGAGTTACAGGGCGGAGCAGGTGCGTGGGATGCCCTCGAC 485
DB 268 GTTCAGCTTCGCCGATGGGTTCATCTCGCGAAGCCGACGATCATCCGGCGGTC 214

RESULT 34
US-10-282-122A-25857/c
; Sequence 25857, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chislen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.

; ORGANISM: Mycobacterium avium
; TYPE: DNA
; ORGANISM: Mycobacterium avium
US-10-282-122A-25857

Query Match      10.7%; Score 53.4; DB 13; Length 1551;
Best Local Similarity 47.0%; Pred. No. 0.00045;
Matches 165; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 125 CGCGGGTTCAGTCTGTCAGTGTACGCGGGGAGGTACAGGATCACAATCTCGCGCAGCCCC 184
DB 1174 CGCGGATCGTGGCGCGGTTCGCGGGGCGAAGCGCAGGATGCGAGCTGCGGGCGA 1115
QY 185 GCGGGAATCTCTCTCTTCGCGCATGATCTCGTGGCGGTGTTCCAGGGGAGAGCAGCG 244
DB 1114 CGCGCACACCGCTCTGTCGTCGAGCGTTCAGTTCAGTTCGCGGTGCGCGCAGAGGT 1055
QY 245 GTPAGTCCACCGGTTCGCGCGGTGAACCGTTCGCGGGTGGCACCGGGATGTGCGTCCCG 304
DB 1054 CGAGTGGCGCGGAGGTCGCAATGGTCCCGGGCGCTGGTTCGCGCACGTGACCCCGCC 995
QY 305 GGTGAGCGCGCCCTCTTGGCGGGTGTGTCGCACACCCAGGAGACAGTCCCGGAC 364
DB 994 CGCGATCACAGCGCATAGCACCGTCTCGGGGCGTACAGTACAGGTGTTGTCGGCG 935
QY 365 CGATGCCGAGAGTTTCGTACGGTGGCGCTTTTCCCGTTCGCGGTACGCGCACCCCC 424
DB 934 AGTGGGAGGAACCGCGCATGCCAGCAGCAGTTCGCGCGCGCGCATGCGGTTCACCGT 875
QY 425 GCTTGGCTTCGGCTTGAGCGAGTTTCAGCAGGGGCGGAGGTCGGTGGCGA 475
DB 874 GGGAGGCCACCGGACACGGGCGAAGCCCGCGGTGCGCGCACCTGGCGCGA 824

RESULT 35
US-10-156-761-3440/C
; Sequence 3440, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3440
; LENGTH: 1884
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1884)
US-10-156-761-3440

Query Match      10.6%; Score 53.2; DB 15; Length 1884;
Best Local Similarity 46.1%; Pred. No. 0.00048;
Matches 213; Conservative 0; Mismatches 248; Indels 1; Gaps 1;

QY 36 TACGCGAGAGGAACACCCCGGTAGTCCGGTACAGGTGCGCGGAGGCGGTACGGCG 95
|||||
```

1313	TACGCCAGAGCCGGTCGGCTCCCGCGCGCTGCGCCGGCGAGACGGGGACGACCCCGTGGACG	1254
96	CTTTCACAGCGTCACGGGGCGGGCGGACACCGGGGGGTACGTCGTCACTGTATACGGGGG	155
1253	GCGGTCCGCG- CGACGGGGCGAGCAGGACGGCGCGTCTCTGTCACGCGCTCCCGGAC	1195
156	GAGGTACAGGATCCACTGTGCGCAGCCCGCGGAGCTCTGTCTCTTCCGCATGATCTC	215
1194	CACGCCGAGGCCAGCGGCTGCAGTGCGCCCGTGAAGCCGTACGCGAGCGTCTGGCANG	1135
216	GTCCGGGTGTTTCCAGCGGAGAGCAGCGGTAGTCCACCGCTGCGGGCGTGAACGGCTC	275
1134	GCGCGCGCGTCCGCGCGAGCGGTGCACATCCGCCACACAGCCGTACGCCAGGCCGT	1075
276	CGGGGTCCGCACCGGGATGTGCGTGCGGGGGTGAGCCGCCCTGTGTTGGCGCGTCTGT	335
1074	CGGTTTCCGGCGGCGTGGTCCGCGAGCGGGCGACCCCGTCGACGCGCGTCACGGCGAC	1015
336	GTGCGACACCCAGGAGACAGGTCCGGACCGATGCCGAGAGTTGTCACGTTGGCGCT	395
1014	CGCGGCGACAAACGGCGCGCCGCTCCCGTCCGACGACCGGGGTGACGCTCGCGTGAAGCC	955
396	TTTCCGCGTCCGCGCGTACGCCACACCCCGTTGCCCTTGGCGCTTGAGCGAGTTACGACG	455
954	CTGGAACGTCGCGGGAGCGCGGCGACCGTGCAGAACCGAGCCGACCGCTCTGCTC	895
456	GCGAGCAGGTGCGGTGCGGATGCCCTCGACGTGGCGGGAA	497
894	GACCCCGGTACCGTGCCTCGTGAAGTGGCGGGGACGGCGGA	853

```

RESULT 36
US-10-239-079-1
; Sequence 1: Application US/10239079
; Publication NO. US2003014846A1
; GENERAL INFORMATION:
; APPLICANT: Merck Patent GmbH
; TITLE OF INVENTION: ANIC-BP-1-ligand
; FILE REFERENCES: ANIC-BP-1-ligand
; CURRENT APPLICATION NUMBER: US/10/239,079
; CURRENT FILING DATE: 2002-09-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2700
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (363)..(2432)
; FEATURE:
; OTHER INFORMATION: Description of ANIC-BP-1 protein ligand
US-10-239-079-1

```

Query Match	10.6%	Score 53.2	DB 15	Length 2700
Best Local Similarity	49.6%	Pred. No. 0.00045		
Matches 136	Conservative 0	Mismatches 138	Indels 0	Gaps 0
QY	218	CGGCGTGTTCACAGGCGAAGACGACGCGTAGTCCACCGGTTCGGGCGTGAACGCGTCG	277	
Db	326	CGGCGGGGCGCGAGCGGCTGCGGTGCGTGGCGCAGATGGGGCGGTGGCACACGGGCCCA	385	
QY	278	GGGTGGCGCACCGGGATGTGCGTCCGCGGGGTGAGCCGCGCCCTGTTCGCGCGGCTCGGT	337	
Db	386	TGTTGGGCTTCGGGTGGCCCGGGCCCGCGAAAGTGTGCGCGGGGCGAGCACCGAGGTGC	445	
QY	338	CGCACACCCAGGAGACACGAGTTCGCGACCGATGCGCGCAGAAAGTTCGTCAACGTTGCGCGTCT	397	
Db	446	AGAGCGCTTGAGCGCGGCGTGGCGGAGCTGCGAGCAGGCGCTTCGCGCTGAAGCAGG	505	
QY	398	TCGCGGTGCGCGCGTACGCCACACCGCTTGGCCCTCGGCGCTTGAGCGAGTTCACGAGG	457	
Db	506	CGGCGCGGCGCGGGCGGTGGCGCGCGCTGGCGCGAGGTCTTCCAACTGTGTGAGGAGG	565	

Qy 458 CGAGCAGTGGTGGCGATGCCCTCGACGTGGC 491  
Db 566 CCTGGTGTCCGGCGGTGGGGCGCGAGGTAGCC 599

```

RESULT 37
US-10-259-165-419/c
; Sequence 419, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaoki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: Patentlist.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 419
; LENGTH: 915
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-165-419

```

Query Match	10.6%	Score	52.8	DB	15	Length	915
Best Local Similarity	49.3%	Pred. No.	0.00068				
Matches	138	Conservative	0	Mismatches	142	Indels	0
Qy	210	GATCTCTCGCGCGTGTTC	CAGGCGAAGACGACGCGCTAGTCC	ACC	CGCTCGGCGCGTGA	269	
Db	306	GATCCCGTACTCGTCC	ACCCGGACTTGATCGCCACG	CAG	TCGTGCGCGACACGATGA	247	
Qy	270	CGGTCGCGGGTGGGCA	CCGGATATGTCGTGTCGGGGGTGAGCGG	CCCTGCTTGGCGG	329		
Db	246	GCAGTCTCTGATCGG	ACGCTGTGGAGCACGAGTCCGGGTTGAT	CCGCTCGGTTGGCGGA	187		
Qy	330	CGTCGTGTGCACACCC	GAGGACCAAGTCCGGACCCGATCCG	CAGAAAGTTCGTCA	CCGT	389	
Db	186	GTGCGTCGGTGCCAG	GATGTCACCCCTCCACCAGATGTTG	CTGCTGTACACCGG	GTG	127	
Qy	390	GGCGCTTTCGCGGTG	CGCGGTACGCCACACCGGTTGCGCTTCG	CGCTTGAGCGATT	449		
Db	126	GATGTTCCACGCGG	GAGTTACACGCTCAGTTCGATATCACC	ACCGTGTCGGAGT	67		
Qy	450	CAGCAGGGCGAGCAG	GTCCGTGCGGATGCGCTCGACGTG	489			
Db	66	CATCACCTCATCAGG	TACCGACCGGTGACTTTGAGCTTG	27			

RESULT 38  
US-10-259-165-79/c  
; Sequence 79, Application US/10259165  
; Publication No. US2003013588A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Tong  
; APPLICANT: Wang, Xun

APPLICANT: Chang, Hur-song  
APPLICANT: Briggs, Steven P.  
APPLICANT: Cooper, Bret  
APPLICANT: Glazebrook, Jane  
APPLICANT: Goff, Stephen A.  
APPLICANT: Katagiri, Fumiaki  
APPLICANT: Kreps, Joel  
APPLICANT: Moughamer, Todd  
APPLICANT: Provart, Nicholas  
APPLICANT: Ricke, Darrell  
TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING  
FILE REFERENCE: 70030-NP  
CURRENT APPLICATION NUMBER: US/10/259,165  
CURRENT FILING DATE: 2002-09-26  
PRIOR APPLICATION NUMBER: US 60/370,620  
PRIOR FILING DATE: 2002-04-04  
PRIOR APPLICATION NUMBER: US 60/368,327  
PRIOR FILING DATE: 2002-03-27  
PRIOR APPLICATION NUMBER: US 60/325,277  
PRIOR FILING DATE: 2001-09-26  
NUMBER OF SEQ ID NOS: 782  
SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta  
SEQ ID NO 79  
LENGTH: 918  
TYPE: DNA  
ORGANISM: Oryza sativa  
US-10-259-165-79

Query Match 10.6%; Score 52.8; DB 15; Length 918;  
Best Local Similarity 49.3%; Pred. No. 0.00068;  
Matches 138; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 210 GATCTCGTGGCGTGGTTCCAGGCGAAGAGCAGCGCGTAGTCCACCGCGTGGGCGTGAA 269  
DB 306 GATCCGCTACTCGTCCACCGGACTTGATCGCACGCGAGTCCGCGCGACACGATGTA 247  
QY 270 CGCGTCCGGGTGCGCACCGGATGTGCTGCGCGGGGTGAGCGCCCTGCTTGGCCGG 329  
DB 246 GCAGTCTCGATCGGACGTGGAGCACGAGTCCCGGGTGTATGCTCGGTGTTGGGCGA 187  
QY 330 CGTGTGCGCACACCCAGGAGACAGGTCCGCGACCGATGCGCGAGAGTTCTGTCACGGT 389  
DB 186 GTGGTCCGTCAGGATCGTCAACCCCTGCACACGATGTTGCTGTACACCGGTTG 127  
QY 390 GCGCTCTTCGCGTGGCGCGTACGCGCACCGGCTTGCCCTCGGCTTGACGAGTT 449  
DB 126 GATGTTCCACGCGCGGAGTTCCACGCGTACGTTGATATCATCACCGGTTCCGAGTG 67  
QY 450 CAGCAGGCGGAGCAGGTGCGTGGGATGCCCTCGACGTCG 489  
DB 66 CATCACTCGATCAGTACCGACCGGCTGACTTGAGCTTG 27

RESULT 39  
US-10-259-165-267/c  
Sequence 267, Application US/10259165  
Publication No. US20030135888A1  
GENERAL INFORMATION:  
APPLICANT: Zhu, Tong  
APPLICANT: Wang, Xun  
APPLICANT: Chang, Hur-song  
APPLICANT: Briggs, Steven P.  
APPLICANT: Cooper, Bret  
APPLICANT: Glazebrook, Jane  
APPLICANT: Goff, Stephen A.  
APPLICANT: Katagiri, Fumiaki  
APPLICANT: Kreps, Joel  
APPLICANT: Moughamer, Todd  
APPLICANT: Provart, Nicholas  
APPLICANT: Ricke, Darrell  
TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING  
FILE REFERENCE: 70030-NP  
CURRENT APPLICATION NUMBER: US/10/259,165

CURRENT FILING DATE: 2002-09-26  
PRIOR APPLICATION NUMBER: US 60/370,620  
PRIOR FILING DATE: 2002-04-04  
PRIOR APPLICATION NUMBER: US 60/368,327  
PRIOR FILING DATE: 2002-03-27  
PRIOR APPLICATION NUMBER: US 60/325,277  
PRIOR FILING DATE: 2001-09-26  
NUMBER OF SEQ ID NOS: 782  
SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta  
SEQ ID NO 267  
LENGTH: 1560  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (310)-(310)  
OTHER INFORMATION: k = guanine or thymine  
US-10-259-165-267

Query Match 10.6%; Score 52.8; DB 15; Length 1560;  
Best Local Similarity 49.3%; Pred. No. 0.00062;  
Matches 138; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 210 GATCTCGTGGCGTGGTTCCAGGCGAAGAGCAGCGCGTAGTCCACCGCGTGGGCGTGAA 269  
DB 847 GATCCGCTACTCGTCCACCGGACTTGATCGCACGCGAGTCCGCGTGGTGGGCGA 788  
QY 270 CGCGTCCGGGTGCGCACCGGATGTGCTGCGCGGGGTGAGCCCGCTTGGCCGG 329  
DB 787 GCAGTCTCGATGCGGAGCGTGGGAGCACGAGTCCCGGTGATCCCGTGGTGGGCGA 728  
QY 330 CGTGTGTCGCACACCCAGGAGACAGGTCCGCGACCGATGCGCGAGAGTTGCTCACCGT 389  
DB 727 GTGCGTGGTGCAGGATCGTCAACCCCTGCACACGAGTTGCTGCTGTACACCGGTTG 568  
QY 390 GCGCTCTTCGCGTGGCGCGTACGCGCACCGCGCTTGCCCTCGGCTTGAGCGAGTT 449  
DB 667 GATGTTCCACGCGCGGAGTTCCACGAGTCAAGTTCGATATCACACCGTGTCCGAGTG 608  
QY 450 CAGCAGGCGGAGCAGGTGCGTGGGATGCCCTCGACGTCG 489  
DB 607 CATCACTCGATCAGTACCGACCGGCTGACTTGAGCTTG 568

RESULT 40  
US-10-156-761-566/c  
Sequence 566, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 566  
LENGTH: 1044  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)-(1044)  
US-10-156-761-566

Query Match 10.5%; Score 52.6; DB 15; Length 1044;  
Best Local Similarity 46.7%; Pred. No. 0.00075;  
Matches 202; Conservative 0; Mismatches 229; Indels 2; Gaps 1;

QY 68 GTAGACGGTGGCGGCAAGCGTACGCGCTTCGACGGTTCACGGCGCGGCGGACACCGG 127  
DB 645 GTCCGCGCCGAAAGCGCCAGGCGCGGCTTCGCGGCGCTGGCGACGGGTGACGGCGATGAT 586  
QY 128 CGGGTCAGTCTC--GTACAGTGTACGCGGGGACGTACAGATCCACTGTCCGCCAGCCCG 185  
DB 585 CGGCGGGCGCGCGCGCGAGGCGGTACGTTGTGCGCGCGCATCCGCGCCACGCC 526  
QY 186 GCGGAACCTCTGCTCTTCGCGATGATCTGCTGCGGCTGCTTCCAGGCGAAGACGAGCGC 245  
DB 525 CCAGACGCGACGAGTGGCGGGGCGCACCCCGCGGTGTGAACGACGCGCGCGGTAGGG 466  
QY 246 GTAGTCCACCGGCTCGGGCGGTGAACCGCTCCGGGCTGCGCACCGGGATGTGCGTCCCGGG 305  
DB 465 GTGGAGACCGGCTCGGGGATGATCGCGGCTGTCGAAGGCGAGGTGTGCGGGATGGG 406  
QY 306 GGTGACCGCGCTGCTGCGCGGCTGCTGTCGACACCCAGGAGACAGGTCCGGACC 365  
DB 405 GATGAGGGTGTCTCGCGCGCGACGCGTGTACTGGGCGCGCGCGCTGTGATGCGATGCC 346  
QY 366 GATGCGCGAGAGTTCGTCAAGTGGCGCTCTTCGCGCGCGGTACCGCGCGCGCGCGCG 425  
DB 345 GCGGCTGCGCATCTGAGTGCAGGGTGGCTTGGCGACGCGCGCGCGCGCGCGAGGT 286  
QY 426 CTTGCCCTCGGCTTGAAGCGAGTTCAGCGGGCGAGCGGTGCGTGGGATGCCCTCGAC 485  
DB 285 CTTGCCGCGCTTCAGGCGTACCGAGTACCGAGTACCGAGTACCGCGCGCGCGCGCGCG 226  
QY 486 GTCCGCGCGGAAC 498  
DB 225 GAGGGTGTGGATC 213

RESULT 41  
US-10-156-761-6641/c  
; Sequence 6641, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: ISHIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6641  
; LENGTH: 1614  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1614)  
US-10-156-761-6641

Query Match 10.5%; Score 52.4; DB 15; Length 1614;  
Best Local Similarity 46.9%; Pred. No. 0.00078;  
Matches 196; Conservative 0; Mismatches 221; Indels 1; Gaps 1;

QY 66 GGTAGACGGTGGCGGCGAAGCGTACGCGCTTCGACGGTTCAGCGCGCGGCGGACACC 125

DB 1493 GCGTAGACGACCTGCGGCCAGATCTCCGGGGCGTC-CTGGTGACCGCTTCGCCGAGCTC 1435  
QY 126 GCGGGTACGCTGTCACGTGTACGGGGGAGAGTACAGGATCCACTGTCCGCGAGCCCG 185  
DB 1434 CGGGTCTGCTTCGCGAGGCGCGGATGTGGAAGGCGAGGACCGTAGTGGCTGGCCAG 1375  
QY 186 GCGGAACCTCTGCTCTTCGCCATGATCTGTCGCGGTGTTCCAGGCGAAGACGAGCGC 245  
DB 1374 GTGCTTGGCGGCTGTCGCGGCGCATGCGGGCGCGGGCGCGCGCGCGCGCGCGCG 1315  
QY 246 GTAGTCCACCGCTGCGGGCGTGAACCGCTCCGGGTCGCGCACCGGGATGTGCGTGGCGGG 305  
DB 1314 GTGGCGACCGCGCGGGGTTGGGACACCGCGAGGGGAGGTTCTTCGCGAGCTGTT 1255  
QY 306 GGTGACCGCGCTGCTTGGCGGGGCTGCTGTCGACACCCAGGAGACAGGTCCGAGCC 365  
DB 1254 GATCGGCTCGAAGTCTTCACCGAGCGATGCGCGCGCGAGCATCCAGCTCTTTCATGAC 1195  
QY 366 GATGCGCGAGAGTTCGTCAAGTGGCGGCTTCGCGCGCGGTTCGCGCGCGGTTCGCGCG 425  
DB 1194 CGTAGCGCGCATGTGGCGGAGGTGTCCTCTTCGCGCGCGGTTCGCGCGCGGTTCGCGCG 1135  
QY 426 CTTGCCCTCGGCTTCAGCGAGTTCAGCGGGCGAGCGGTGCGTGGGATGCCCTCG 483  
DB 1134 GCGGCGCTCGGTGACGACCGCTTCGCGCTTCGAGGTGTGCGCGCGGACCGCG 1077

RESULT 42  
US-10-156-761-6347/c  
; Sequence 6347, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6347  
; LENGTH: 1182  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1182)  
US-10-156-761-6347

Query Match 10.4%; Score 52.2; DB 15; Length 1182;  
Best Local Similarity 48.3%; Pred. No. 0.00091;  
Matches 216; Conservative 0; Mismatches 218; Indels 13; Gaps 2;

QY 53 CCGCGGTAGTTCGCGGTAGCGGTGGCGGCGAAGCGGTACGCGCTTCGACGGTTCAGCGG 112  
DB 810 CACGAGACGTACGCGGTGCTGCGGCTCTTCGCGCGCACAGGGTGGCGGCGAGTTCGCG 751  
QY 113 GCGGCGGACACCGCGCGGTGAGTTCGCTGCTGTCAGCGGTGACCGGGGACGTACAGGATCCACT 172  
DB 750 GCGCGCGAGTTCG-----TCGTACCGCGCACACAGGAGGTGTGCTGCGAGTTCGA 701  
QY 173 GTCGCGAGCGCGCGGAGTCTCTGCTTCGCGCATGATCTCTGCGCGGTGCTTCAGG 232  
DB 700 TGTGTTGGCGCGCGAGCGTTCGCTTCGATGTGCGGTGACGCTCTCGCGCGGGTGGAGC 641



QY 233 CGAAGACGACGGCGTAGTCACCGCGTCGGCGTGAAACGGCTCCGGGTGCGACCGGA 292  
DB 640 CGTCGAGCGAGGCGCGTGGCGACGCGCTGCGCGAGCGGATTTGGTGAAGACCGCGTCGA 581  
QY 293 TGTGCGTCCGCGGGGTGAGCGCGGCCCTGCTTGGCGCGCGTCTGCGCACACCGCAGGAGA 352  
DB 580 TCAGTTCGAGCGCGTTTCAGCGCGC--GCTTGGCCCGCCAGGCGCGACCGCGTAGTACG 524  
QY 353 CCAGGTCGCGACCGATGCGCGAGAGTTGTCAGGTTGGCGCTCTTCGCGGTGCGCGCGT 412  
DB 523 CTTGTCGCGCGCGCCAGGCGCGGCCCGCCCGATGCGCGCGCACCGTTCGCGAGCGCGCCCT 464  
QY 413 ACGCCACACCGCTTCCCTCGGCTTGAGCGAGTTTCAGAGGCGCGACGAGTCCGTTGC 472  
DB 463 CGGGGATCAGGTTTCGAGTCCGCGGACGCGCGCCGAGCCAGCGACTGGAAGCCGATCT 404  
QY 473 GGATGCGCTCGACGTCGCGGCGGAACC 499  
DB 403 GGCAGACGACTCCGCGCGCGCGACCC 377

RESULT 43  
US-10-282-122A-25426  
; Sequence 25426, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25426  
; LENGTH: 1200  
; TYPE: DNA  
; ORGANISM: Mycobacterium avium  
US-10-282-122A-25426

Query Match 10.4%; Score 52.2; DB 13; Length 1200;

Best Local Similarity 52.0%; Pred. No. 0.00091;  
Matches 117; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
QY 266 TGAACGCGTCGCGGGTGCACACCGGATGTGCTGCGCGGGGTGAGCGGCCCTGCTTGG 325  
DB 422 TGCTCAAGCGCGGGGTGTTTCACGACATCTCCGCGGGGTGATGCTGATCCCGGGCGCG 481  
QY 326 CGGGGTGCTGTCGACACACCGAGGACCGAGTCCGACCGATGCCGAGAGTTCTGTCA 385  
DB 482 CGACATCGCGCGCGCGCTGCTGCGCTTTCGAGGCGATGCTGGGCTACCGCGGCA 541  
QY 386 CGTGGCGCTCTTTCGCGCTGCGCGGTACGCGCACACCGCTTGCCTTGGCGCTTGAAGC 445  
DB 542 AGGAATCGACGCGCGCGCTGCGCGCGCACCTGGGGGTCAACGCGCTCGACCGCGTACCG 601  
QY 446 AGTTTCAGCAGCGCGGAGGAGTGGTGGGATGCCCTGACGCTCG 490  
DB 602 TCGCGAAGTGCGCGTCTGCTGCTCAGCAACTGGCGCCG 646

RESULT 44  
US-10-156-761-5628/c  
; Sequence 5628, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 5628  
; LENGTH: 1425  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1425)  
US-10-156-761-5628

Query Match 10.4%; Score 52; DB 15; Length 1425;  
Best Local Similarity 48.6%; Pred. No. 0.00099;  
Matches 142; Conservative 0; Mismatches 150; Indels 0; Gaps 0;  
QY 194 CTGTCTCTTCGCGATGATCTGTCGCGGTGTTCCAGCGGAGAGCGCGGTAGTCA 253  
DB 1363 CCAGGAGCGTCTCACCAGCGGTCCCGATGCCCGCGGACACCGGGCCAGGAAGA 1304  
QY 254 CGCGTCCGGCGCTGAACCGTCCGGGTGCGCACCGGATGTGCGTCCGGGTGAGCC 313  
DB 1303 GCGAGCGGAGCGTTCGCGACCGCCAGCGCCCGACTGCTGGGTGATCATCAGC 1244  
QY 314 GGCCTTGTGCGCGCGTGTGTCGCAACCCAGGAGACCGAGTCCGAGCCGATGCCGC 373  
DB 1243 CGCGCGCCACGCGCGCGGTGCGCGGCACCTCGGACAGGATGATCCGAGAGAGACGG 1184  
QY 374 AGAAGTTCGTCACGGTGGCGCTTTCGCGGTGCGCGGTAGCCACCCCGCTTCCCT 433  
DB 1183 GCAGTGGAGCGCTTGGCGCGCGCCCGCGCGCCAGCCCGCGAGCTCCAGCAGC 1124  
QY 434 CGGCCTTGAGCGAGTTTCAGCAGCGGAGCAGGTGCTGGCGGTGCGCTCGAC 485  
DB 1123 CCAGGTCCGGCCAGGACCGCGAGCGCGGCGGACGATCAGGGTCAACCCGAC 1072

APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(5313) B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128

QY 310 AGCGGCGCTGCTGGCGGGTGTGTCGCACACACCCAGGACAGGTCGCGACCGATG 369  
Db 554 CGGCCCAGCAGGCTGACCGCGGGGTGCGCGGCCCGGACATGCGGATCCGACACG 495  
QY 370 CCGCAGAAGTTCGTCACGCTGGCGCTCTTCGCGGTGCGCGGACACCCCGCTTG 429  
Db 494 CCGCGCTCGACGACGCGCTCGCGGGCGGCTCCAGCAGCGCGGCTCGCGCGAGCTTG 435  
QY 430 CCCTCGGCTTGACGAGTTCACAGGCGGAGCAGGTCGCGTGGGATGCCCTCGAGCTCG 489  
Db 434 GCCTTGGCGTAGTGGTTCGCGCGGGCGGCGTCCGCGTCCGCGGTGCTCTCGCGGGC 375  
QY 490 GCGG 493  
Db 374 GCGG 371

## RESULT 48

US-10-282-122A-11252  
; Sequence 11252, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; PRIOR FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 11252  
; LENGTH: 2010  
; TYPE: DNA  
; ORGANISM: Burkholderia cepacia  
US-10-282-122A-11252

Query Match 10.2%; Score 51.2; DB 13; Length 2010;  
Best Local Similarity 53.3%; Pred. No. 0.0015;  
Matches 130; Conservative 0; Mismatches 113; Indels 1; Gaps 1;  
QY 24 AACGAGTCGAGATACGCGGAGGAGAACACCCCGCGGTAGTCCGGGTAGACGGTGGCGCG 83

Db 1612 AACCTGCATACCGACGGCTATACGAAACCGGCGCGCGCTGCGCGCGGGC 1671  
QY 84 AAGGGTACGGCGCTTCGACCGTACGCGGGGGGGGACACCGCGGGTACGCTCGTCAC 143  
Db 1672 GAGACG-ACCCACGTCGCGCTTCTCGACGCTCGGCTCGGTGCGCGGTGCGAGCTCGGCTC 1730  
QY 144 GTGTACGCGGGGAGCGTACAGGATCCACTGTCCGCGACGCGCGGGAACCTCTGCTCTCTT 203  
Db 1731 GATCGCAGCGCGCAGTTTCAACCGCGCGGAGCGTGGCTGGCGACATGCTTCGGCAA 1790  
QY 204 GCCCATGATCTGTCGCGGTGTTCCAGGCGGAGAGACGCGGTAGTCCACCGCGTGGG 263  
Db 1791 CGTGGCGCGCTGCTGCGGCTTTCACGTTTCGCAATGGCGGCGACGTCGTTCCAGGTGTGGG 1850  
QY 264 CGTG 267  
Db 1851 CGTG 1854

## RESULT 49

US-09-922-683-3/c  
; Sequence 3, Application US/09922683  
; Publication No. US20020192793A1  
; GENERAL INFORMATION:  
; APPLICANT: DECKER, Heinrich  
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR  
; PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS  
; GLA.O AND THEIR USE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/922,683  
; FILING DATE: 07-Aug-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/194,905  
; FILING DATE: 1999-12-01  
; APPLICATION NUMBER: DE 19622783.6  
; FILING DATE: 07-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granados, Patricia D.  
; REGISTRATION NUMBER: 33,683  
; REFERENCE/DOCKET NUMBER: 026083/0193  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 546 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-922-683-3

Query Match 10.1%; Score 50.6; DB 9; Length 546;  
Best Local Similarity 49.8%; Pred. No. 0.0026;  
Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;  
QY 239 GCAGCGCGTAGTCCACCGCGTGGCGGTGAACGCGTCCGGGTGCGCACCGGATGTGCG 298

Db 383 CGAGGAGCGGTACACTCGTGGTGGAGACGTGACGAGGGTGGCGCACCGTGGCGGA 324  
Qy 299 TCCCGGGGTGAGCGCGCCCTCTTGGCGGGTCTGTCGCACACCCAGGAGACCAGGT 358  
Db 323 GCGCGCGTCGAGCAGGACCTGGTGGCCAGCAGTGTGGTGGGTGAAGGACCGCTGT 264  
Qy 359 CGGACCGGATGCGCAGAAAGTTGTCACGGTGGCGCTCTTCGCCGTGCGCGGTACGCCA 418  
Db 263 CGGTGATGAGCGGTGACGAGTGGCACTCGGCCGGAAGTGACAGATGTCTGTGCGCGG 204  
Qy 419 CCACCGCTTGCCTCGGCTTGGCGGAGTTGACAGAGGGGAGCAGGTGCGTGGGATGC 478  
Db 203 CGGCCAGCGTGTGACGAGCGGGTGTGCGACACGTGCGCCCTGGACGAAGGTGAGGCGG 144  
Qy 479 CCTCGACGTGCGCGCG 495  
Db 143 GATGTCACGACCGCG 127

## RESULT 50

US-10-156-761-2762/c  
; Sequence 2762, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2762  
; LENGTH: 1260  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1260)  
US-10-156-761-2762

Query Match 10.1%; Score 50.6; DB 15; Length 1260;  
Best Local Similarity 47.9%; Pred. No. 0.0022;  
Matches 146; Conservative 0; Mismatches 159; Indels 0; Gaps 0;  
Qy 188 GGAATCTCTCTCTTCGCCATGATCTGTCGGCGTGTTCAGGGGAGAGAGCGGT 247  
Db 592 GGTAAGCGCGGACCTCGTCCGCGATCGGGGGAACCGCGGAAAGTCCAACTGGCGCGGT 533  
Qy 248 AGTCCACCGCGTGGGGGTGAACCGCTCCGGGTGGCGACCGGGATGTGCTGCCGGGG 307  
Db 532 ACGCGACAGCGCGGCCACGATCAGCTTCGGCTGGACCTCTTGGCGACCTTCTGACCT 473  
Qy 308 TGAGCGCGCCCTGCTTGGCGGGTGTGTCGCACACCCAGGAGACAGGTCCGGACCGA 367  
Db 472 CGGCAATGTCCACCTGGCGGGTGTCTGTCACGTGTGTCACGACACGCTTGTACAGCT 413  
Qy 368 TGCGCAGAAAGTTGTCACGCTGCGGTCTTCGCCGTGGCGCGGTACGCCACCCCGGT 427  
Db 412 TGCGGAGAAAGTTGATCTTCATGCCGTGGGTGAGTCCCGCGCTGGCGAGGTTTCAGAC 353  
Qy 428 TGCCCTCGGCTTGGAGGAGTTGACAGGGCGGAGCGAGGTGCGTGCAGATGCCCTCGACGT 487  
Db 352 CCATGATCGTGTGCCCGGGTGTGACAGCGGAAACATGGCGCGCGGTGGCTGGCGCG 293

Qy 488 CGGCG 492  
Db 292 CCGAG 288

Search completed: June 27, 2004, 21:02:34  
Job time : 360.804 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 18:07:12 ; Search time 2028.43 Seconds  
(without alignments)  
7360.905 Million cell updates/sec

Title: US-09-758-759-1\_COPY\_1\_500

Perfect score: 500

Sequence: 1 ggtaccgaccgtgtccgg.....tcgacgtcggggcgacact 500

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database :

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estlin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_man:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_pbg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	63.2	12.6	925	CNS0091P	AL053013 Drosophila
c 2	61.8	12.4	982	BX415111	BX415111 BX415111
c 3	61.6	12.3	982	BX415111	BX415111 BX415111
c 4	61.4	12.3	623	BQ295055	BQ295055 WHE2857_G

RESULT 1  
CNS0091P/c

LOCUS  
DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BAC19D16 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

AL053013  
GI:4934461

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 925)

Genoscope.

Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage :

BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

ALIGNMENTS

c 5	61.4	12.3	935	29	CNS006XK	AL066051 Drosophil
c 6	61	12.2	925	29	CNS0091P	AL053013 Drosophil
c 7	59.8	12.0	512	9	AV623523	AV623523 AV623523
c 8	59.8	12.0	512	9	AV625284	AV625284 AV625284
c 9	59.8	12.0	532	9	AV619788	AV619788 AV619788
c 10	59.8	12.0	599	12	BM002630	BM002630 1031105G0
c 11	59.8	12.0	786	10	BF860172	BF860172 963015G07
c 12	59.4	11.9	1085	29	CNS035KP	AL228850 Tetradon
c 13	59	11.8	932	29	CNS0072Q	AL066742 Drosophil
c 14	58.8	11.8	750	10	BF860173	BF860173 963015G07
c 15	58.6	11.7	918	29	AG043040	AG043040 Pan trogl
c 16	58.6	11.7	1101	14	CK207202	CK207202 FOAS01882
c 17	58.2	11.6	581	14	CA622308	CA622308 wln.pk00
c 18	58	11.6	932	29	CNS0072Q	AL066742 Drosophil
c 19	58	11.6	935	29	CNS006XK	AL066051 Drosophil
c 20	57	11.4	588	28	BZ365371	BZ365371 ie15a07.g
c 21	57	11.4	755	28	CC334644	CC334644 OQAM79TH
c 22	57	11.4	772	29	CG275728	CG275728 OXFI14TH
c 23	57	11.4	1538	29	AG030607	AG030607 Pan trogl
c 24	56.8	11.4	948	13	BQ646304	BQ646304 AGENCOURT
c 25	55.8	11.2	503	9	AL820302	AL820302 AL820302
c 26	55.8	11.2	779	29	CG324766	CG324766 QGWB31TV
c 27	55.8	11.1	789	29	CC662995	CC662995 QGUJ523TH
c 28	55.6	11.1	559	13	BX384405	BX384405 BX384405
c 29	55.6	11.1	563	14	CB678034	CB678034 OSUNel5J
c 30	55.4	11.1	986	29	AG074680	AG074680 Pan trogl
c 31	55.2	11.0	456	13	BQ080807	BQ080807 1030007C0
c 32	55.2	11.0	500	10	BF630057	BF630057 HVSMB5000
c 33	55.2	11.0	828	13	BX425797	BX425797 BX425797
c 34	55	11.0	736	14	CB678003	CB678003 OSUNel151
c 35	55	11.0	774	14	CB651766	CB651766 OSUNel16N
c 36	55	11.0	777	14	CB672283	CB672283 OSUNel06D
c 37	54.8	11.0	810	14	CB671209	CB671209 OSUNel04K
c 38	54.8	11.0	924	13	BX442207	BX442207 BX442207
c 39	54.8	11.0	1201	9	AL538546	AL538546 AL538546
c 40	54.4	10.9	605	13	CA075606	CA075606 SCJLAM106
c 41	54.4	10.9	938	29	CG262283	CG262283 QG2CH86TH
c 42	54.2	10.8	905	28	CC443238	CC443238 PUHEM44TD
c 43	54.2	10.8	916	29	CG268544	CG268544 OXAZ444TV
c 44	54	10.8	596	14	CA758758	CA758758 BR0300040
c 45	54	10.8	866	29	CG447961	CG447961 OGA9A78TH
c 46	53.8	10.8	851	29	CG338271	CG338271 OGOCG91TH
c 47	53.8	10.8	1798	29	AG171124	AG171124 Pan trogl
c 48	53.4	10.7	665	14	CB659824	CB659824 OSJNed01B
c 49	53.2	10.6	828	29	CG062012	CG062012 PUJFJ27TD
c 50	53	10.6	581	28	BZ359621	BZ359621 id79c11.b

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Ooeogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RP1-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source

```
1..925
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RP1-98"
/notes="end : TET3"
```

## ORIGIN

```
Query Match      12.6%; Score 63.2; DB 29; Length 925;
Best Local Similarity 13.7%; Pred. No. 3.5;
Matches 45; Conservative 161; Mismatches 122; Indels 0; Gaps 0;

QY 168 CCACGTGTCGCGAGCGCGGGAACCTCGTCTCTCGCATGATCTCGTGGCGGT 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 921 CSCSCSBSBSCSSSSMSSTSSNSBSCSSBSSTSSMSBSBSBSBSBSBSST 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 228 CCAGCGAAGACGACGCTAGTCACCGCTCGGCGGTGAACCGCTCGGCGGTGCCAC 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 861 SSACVKNASSSCCGCGMABCMCSSSSSCGASARGYKVRASGAGKRGCGSGGAS 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 288 CGGATGTCGTGCGGCGGCTGAGCCGCTCTGCTGCGGCTCTGTCGACACCCA 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 801 ASHSSSSACBSSSSCSASWSASSSSSASRSRSGGAGGASRSRSSSSSSASAG 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 348 GGAGACACAGTCCGACCGATCCGACAGATTCTGTCACGCTGCGCTCTTCGCGCTGC 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 741 SVVSSASSSSSSSSSVSCSVASMSCSBSSSSSASASSSSSSSASCSACSCCTTWS 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 408 CCGCTACCCACACCGCTTGCCTCGGCTTCGAGCGAGTTACAGAGGCGACAGTC 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 681 CECSTASMSAARSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 468 GTGCGGATGCCCTGACGTGCGCGGCG 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 621 GGGSGSVSAGSMSSSVSSSGRSGSG 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 2

```
EX415111
LOCUS      EX415111 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004Y120
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION  EX415111
VERSION     EX415111.1 GI:30765470
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 982)
AUTHORS   Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
```

Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0CAP004B10P1.

## FEATURES

source

```
1..982
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP004Y120"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
```

## ORIGIN

```
Query Match      12.4%; Score 61.8; DB 13; Length 982;
Best Local Similarity 28.9%; Pred. No. 5.7;
Matches 118; Conservative 100; Mismatches 191; Indels 0; Gaps 0;

QY 27 GAGTCGAGATACGGCGAGAGAAACACCCCGGTAGTCCGGTAGACGGTGGCGCGAAG 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126 GGGSSSGVGVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 87 GCGTACGCGCTTCGACCGGTACGCGGCGGCGGACACCGCGGGTTCAGTCTGTCACGTG 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 186 SGGKCCCGKCCCCCGGGGGGGGGCCSSSSSSSSSSSSSSSSSSSSSSSSSSSS 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 147 TAGCGGGGAGCTACAGATCCACTGTCCGCACGCCGCGGGAACCTCTGCTCTTCCG 206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 246 CCCCCCCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 207 CATGATCTCTGTCGGCTGGTTCCAGCGAAGACGACGCGTAGTCCACCGGTGCGGCGT 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 306 GGGCGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 267 GAACGCTCGGGGTGCGCACCGGATGTGCTGCGGGGGGTGACCGCGCTGCTTGC 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 366 SSSCCGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 327 CGCGCTGCTGTGCGCACACCCAGAGACGAGTCCCGACCGATGCCGAGAGATTCTGCAC 386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 426 GSGSGSGSGSVGGGSSCGSCGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 387 GTTGGGCTCTTCGCGCTGCGCCGTACGCCACACCGCTTGCCTCG 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 486 GGGGGGGCGSSGSGGGGGGGGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 3

```
EX415111/c
LOCUS      EX415111 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004Y120
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION  EX415111
VERSION     EX415111.1 GI:30765470
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 982)
AUTHORS   Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
```

```

FEATURES
source
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSOCAP004BE10QPI.
Location/Qualifiers
1. .982
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOCAP004YI20"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(4T) primer. Five prime end enriched, into
double-strand cDNA was digested with Not I and cloned, into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

```

```

ORIGIN      Query Match    12.3%; Score 61.6; DB 13; Length 982;
            Best Local Similarity 27.1%; Pred. No. 6.1;
            Matches          92; Conservative 100; Mismatches 148; Indels 0; Gaps 0;

QY   81  GCGAAGCGGTACGCGCCTTCGAGCTCAGCGGGCGGCACACCGCGGGTGAGTCCGT 140
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   451 SCGGCGSCGCCCBCCCCSSCCSSCCCSCSGGGCGGGCCSSSSSKSCCCCSSS 392
           :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY   141 CAGCTGTACGCGGGGACAGTTACAGGATCCACTGTCCGCAGCCCGCGGAACTCCTGTCTC 200
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   391 CCCCSSSCCCCCCSSGSSSCGSSSSSCCCSCSGSGSCSGGCSHGSSGSSCCCCCGGGSC 332
           :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY   201 CTTCGCCCATGATCTGTCGCGCTGTTCTCAAGGCGAAGACGACGCGTAGTCACCACCGTCTC 260
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   331 SCSSSSSSCCSCCSGGSSCGCCCCSCCSCGSCGSCGSCGSSSSSCSGSSGSS 272
           :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY   261 GGCGGTGMA CGCTCCGGG GTGCGCACCGGAGATGCGTCCCGGGGTGAGCGGCCCTG 320
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   271 GGCGGCCCCSCGGGGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 212
           :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY   321 CTTCGGCGGCGTGTGTCGCACACCCAGGAGAACAGTCCCGGACCGATGCCGCAGAAGTT 380
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   211 CCCCS CCGGGGGGMCWGGMWCGSSSGGSGSCSBSSGSGSCCCCCCRGRGGGSSS 152
           :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY   381 GTTACAGTGGCGCTCTTTCGCCGTTCGGCGCGCTGAGCCACC 420
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   151 SSSSSSSSSSSSBGBCSBSSSCCCSCCCSCCCSCCCSCCCSCCCSCCCV 112
           :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

RESULT 4  
 BQ295055/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticeae; Triticum.  
 1 (bases 1 to 623)  
 Anderson,O.D., Chao,S., Chin,A., Close,T.J., Crossman,C.,  
 Gustafson,P., Lazo,G.R., Pham,J., Rausch,C.J., Ross,K., Wilson,C.,  
 and Woo,J.  
 the structure and function of the expressed portion of the wheat  
 genomes - Unstressed root tip cDNA library  
 Unpublished (2002)  
 Contact: Olin Anderson  
 US Department of Agriculture, Agriculture Research Service, Pacific  
 West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 5105595773  
 Fax: 5105595818  
 Email: candersh@pw.usda.gov

```

Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.
Location/Qualifiers
1. 623
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE2857.G05.N09"
/tissue_type="Root tip at 1.0 to 1.5 mm stage"
/dev_stage="Four-day old seedling"
/lab_host="E. coli SOLR"
/clone_lib="Wheat unstressured root tip cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown under;
hydroponic conditions for four days. Root tips were
excised and snap frozen (Ross and Gustafson) and total RNA
was prepared at University of Missouri, Columbia. Poly(A)
RNA was purified, a cDNA library was made, and the cDNA
clones were in vivo excised to give pBluescript
SK(-) phagemids in the TJ Clouse lab (Chin and Clouse) at the
University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."

```

```

ORIGIN                                Anderson lab (all other authors).

Query Match      12.3%; Score 61.4; DB 13; Length 623;
Best Local Similarity 49.8%; Pred. No. 6.3;
Matches 155; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY      189 GAATCTCTGTCCTTGC CATGATCTCGTCGGCGGTGGTTCCAGGCGAAGACGACGCCGTA 248
       |||||
Db      357 GAGCTCGGTCTCCTCGGACGACCCCTCGGGGTCCATGTGGGCGCTCGACCTCGGCGGC 298
       |||||
QY      249 GTCCACCGGTGGGGCGTGAAACGCGTCCGGGTGCGCACCGGATGTGCTGCGGGGGT 308
       |||||
Db      297 CCGCA CGGCGTCCGGCGCTCGCAGTAGCGAAGAGGGCCGGGCCGTAGAGCGGAGCG 238
       |||||
QY      309 GAGCGCGCCCTGTTGGCCGGGCTCGTGTGCGACACCAGGAGACCAAGGTCCCGACCGAT 368
       |||||
Db      237 CGCGGGGATCCCGCGCGCGCCATGCGGGCGGACGACGGAGAAGCCATGCGCGGTCCCC 178
       |||||
QY      369 GCGCAGAGATTGTCACGGTGGCGCTCTTCGGCGGTGCGGCGGTAGGCCACCAACCGCTT 428
       |||||
Db      177 GACGCGCGCGCGAGCGCGCGGAGGTGTGTGAGCTGGGCTCTGTGGGCTCCACGCGCTG 118
       |||||
QY      429 GCCTTCGGCTTTGAGCGAGTTTCAGCAGGGCGACAGTTCGGTGGGATGCGCTTCGACGTC 488
       |||||
Db      117 CGCCTCATGGCGCGGAAGATCTCGAAGCGCGGCGGCGCTCGGCGTCTCTGGCGCG 58
       |||||
QY      489 GGCGCGCAACC 499
       |||||
Db      57 CGCGTTGGAGC 47
       |||||

```

RESULT 5	
CNS006XX/c	
LOCUS	
DEFINITION	CNS006XX linear GSS 03-JUN-1999 Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14N09 of RCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	ALJ066051
VERSION	ALJ066051.1 GI:4945019
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 935) Genoscope.
REFERENCE	
AUTHORS	
TITLE	Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr]  
- Web : www.genoscope.cns.fr

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library](http://www.fruitfly.org/The%20BDGP/Drosophila%20melanogaster/BAC%20library.html) was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pictet de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
Location/Qualifiers  
1..935  
    `/organism="Drosophila melanogaster"`  
    `/mol_type="genomic DNA"`  
    `/db_xref="taxon:7227"`  
    `/clone="BACR14N09"`  
    `/clone.lib="RPCL-98"`  
    `/note="end : T7"`

ORIGIN  
  
Query Match                  12.3%; Score 61.4; DB 29; Length 935;  
Best Local Similarity       25.3%; Pred.No. 6.6;  
Matches 108; Conservative 122; Mismatches 197; Indels 0; Gaps 0;

QY     66 GGGTAGACGGTGGCGGCGAAGCGGTACGCGCCCTTCGACGGTGACGGGCGGCGGCACACC 125  
      |||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
DB     935 GSGCGGGGGGGSSGSGCGGCGSGSGSCCCCSCGCCSCGCCSCCSCCSSCCCC 876  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
QY     126 GCGCGGTTCAGCTCGTACGTGACGCGGGGACGTACAGGATCCAATGTCGCCACGCC 185  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
DB     875 SSGCGCGCGCGSGSGSCCGCGCGSSSGCGGCCCGCGCGCGCGSCCGSCCS 816  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
QY     186 GCGGAATCTTCCTTCGTCATGATCTCGTGGCGGTGTTCCAGGCGAAGACGAGCGC 245  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
DB     815 CCG 756  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
QY     246 GTAGTCCACCGCGTTCGGCGGTAAACCGGTTCGGGGTGCGCACGGGATGTGGTCCCGG 305  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
DB     755 GSGCSSSSGSGSGSGSGSGSGSCSGSCSGSGSGSGSGSGSGSGSGSGSGSGSG 696  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
QY     306 GGTGAGCCGCGCTTGTGGCGCGGTGTGTGTGCAACCCAGAGACAGATGTCGGACC 365  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
DB     695 GCGSGCGSGCG 636  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
QY     366 GATCGCGCAGAATTTCGTACGGTGGCGCTCTTCGCGGTGCGCGCTACGCCACACCG 425  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
DB     635 GCCCG 576  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
QY     426 CTTCGCCCTCGGCTTAGCGAGTTTCAGACGGGCGAGCAGTCCGTCGATCCCTCGAC 485  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
DB     575 GSASGCGCGMGCRAGSGKMGAGSGRCGGACSGSGGSBSSTKKGASCACSCSAY 516  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
QY     486 GTCCGGC 492  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
DB     515 GKSGSGC 509  
      |||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:

**KEYWORDS**  
Drosophila melanogaster (fruit fly)

**SOURCE ORGANISM**  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 925)  
Genoscope.

**REFERENCE**

**AUTHORS**

**TITLE**

**JOURNAL**

**COMMENT**  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrf@genoscope.cns.fr ;  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see http://www.fruitfly.org The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammeter in Peter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
PcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
pi and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.  
Location/Qualifiers  
1..925  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR19D16"  
/clone\_lib="RPCI-98"  
/note="end : TEf3"

**FEATURES**  
source

**ORIGIN**

Query Match	12.2%;	Score 61;	DB 29;	Length 925;
Best Local Similarity	13.9%;	Pred. No. 7.5;		
Matches	45;	Conservative 157;	Mismatches 122;	Indels 0; Gaps 0;

```

QY   172  TGTCCCGCCGCGGGGAACCTCCTGCTTCCTTGCATGATCTGTCGGCGTGTTCCAG 231
Db    562  KGCSSGGBSCCSSCSGSSSCBCCCCCCSGSSGCCSSBSBKSTSBSCSCC 621

QY   232  GCAGAAGACGACGCGTAGTCAACCGCTGGCGGTGAACGCTCCGGGGTGCACACGGG 291
Db    622  SSKSVGCTSCSSSSSCSSSSSTTSBSTSSTSKSSSGSSSSSSSYTTKSTSASGS 681

QY   292  ATGTGGTGC CGGGGTGACCGCGCTGTTGGCGGCGTCTGTGCGCACACCACGAG 351
Db    682  SWAGGGSGTGTTSSSSSSTTSSSSVSSGSKSTBSGGBSSGSSSSSTSSBS 741

QY   352  ACCAGTCCGACCGATGCGCGAAGATTGTCACGCTGGCGTCTTTCGCGCTGCGCCG 411
Db    742  CTSTSSSSSSSVSSSTCCTCCCSYSYSTSSSTSSSSTSMGSTSGSSSSVGTSSSDST 801

QY   412  TAGCCACACACCGCTGCGCTCGGCTTCAGCGAGTTCAAGACGGGCGACAGTCCGTG 471
Db    802  STCCSCCCYMCTCSBYMBCYNTSCTGCGSSSSSGKGVTCKCGCGCSSTNGMBGTSS 861

QY   472  CGGATGCCCTCGACGTGGCGGCG 495
Db    862  ACSSSSSSCSSSVSSSSSKSSASS 885

```

**RESULT 7**  
AV623523/c LOCUS AV623523 512 bp mRNA linear EST 15-DEC-2000  
DEFINITION AV623523 Chlamydomonas reinhardtii 5' to 0.04% CO2 Chlamydomonas  
reinhardtii cDNA clone LC065b05\_r 5', mRNA sequence.  
ACCESSION AV623523  
VERSION AV623523.1 GI:10772700



**KEYWORDS**  
**SOURCE** Chlamydomonas reinhardtii  
**ORGANISM** Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

**REFERENCE**  
**AUTHORS** Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y., and Tabata, S.  
**TITLE** Generation of expressed sequence tags from low-CO<sub>2</sub> and high-CO<sub>2</sub> adapted cells of Chlamydomonas reinhardtii  
**JOURNAL** DNA Res. 7 (5), 305-307 (2000)  
**MEDLINE** 20539644  
**PUBMED** 11089912  
**COMMENT** Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

**FEATURES**  
 source  
 1..512  
 Location/Qualifiers  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone="LC090d07\_r"  
 /clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO<sub>2</sub>"  
 /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"

**ORIGIN**  
 Query Match 12.0%; Score 59.8; DB 9; Length 512;  
 Best Local Similarity 48.7%; Pred. No. 11;  
 Matches 163; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 124 CCGCGGGGTACGTCCTCGTACGCGGGGACGTACAGATCCACTGTCGCCAGCC 183  
 Db 373 CCCAGGAGGAGTCTATGNGGGGATGGCAGCGCTGCTGATGCACGCGGCCAGC 314  
 QY 184 CGCGGAACTCTGCTCTTGGCCATGATCTGTCGGCGTGTTCACGCGGAAGACGAC 243  
 Db 313 CGCGCGCAGACGCCCTCGGACGCGGTGAGTCTCTTGGCGAAGTCGGGGTCCACGACG 254  
 QY 244 GCGTAGTCCACCGCGTCGGCGTGAACGCGTCCGGGTGCGCACCGGGATGTCGTCGG 303  
 Db 253 GAGGCGAGTCTGGGGTGGCTCTGATAGGCGACGCGGATCGGTCAGGAACTTCGCGCGG 194  
 QY 304 GGGGTGAGCGCGCCCTGCTTGGCCGCGTGTGTCGACACCCAGGAGACGAGTCCGGA 363  
 Db 193 ATGATGAGCAGCACCTTCAGATGCGCGCAGCGCCCGAGGTTCGACACTTCATC 134  
 QY 364 CGATCGCGCAGAGTTCGTCAGGTGGCGCTCTTCGCGTCCGCGTACGCCACCCACC 423  
 Db 133 TCACGCTCTTGGCCTTGATGTTTCATGCCCTGCGGTACGAGATCTTGTGGCG 74  
 QY 424 CGCTTGCCTCGGCTTGACGCGAGTTTCAGCAGGCG 458  
 Db 73 TACAGCGCGCGGACGCTGTTGATCAGTGGCG 39

**RESULT 8**  
**AV625284/c**  
**LOCUS** AV625284 512 bp mRNA linear EST 15-DEC-2000  
**DEFINITION** AV625284 Chlamydomonas reinhardtii 5% to 0.04% CO<sub>2</sub> Chlamydomonas reinhardtii cDNA clone LC090d07\_r 5', mRNA sequence.  
**ACCESSION** AV625284  
**VERSION** AV625284.1 GI:10774461  
**KEYWORDS** EST.  
**SOURCE** Chlamydomonas reinhardtii  
**ORGANISM** Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.  
 1 (bases 1 to 512)  
 Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y., and Tabata, S.  
 Generation of expressed sequence tags from low-CO<sub>2</sub> and high-CO<sub>2</sub> adapted cells of Chlamydomonas reinhardtii  
 DNA Res. 7 (5), 305-307 (2000)  
 11089912  
 Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

**FEATURES**  
 source  
 1..512  
 Location/Qualifiers  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone="LC090d07\_r"  
 /clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO<sub>2</sub>"  
 /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"

**ORIGIN**  
 Query Match 12.0%; Score 59.8; DB 9; Length 512;  
 Best Local Similarity 48.7%; Pred. No. 11;  
 Matches 163; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 124 CCGCGGGGTACGTCCTCGTACGCGGGGACGTACAGATCCACTGTCGCCAGCC 183  
 Db 506 CCCAGGAGGAGTCTATGNGGGGATGGCAGCGCTGCTGATGCACGCGGCCAGC 447  
 QY 184 CGCGGAACTCTGCTCTTGGCCATGATCTGTCGGCGTGTTCACGCGGAAGACGAC 243  
 Db 446 CGCGCGCAGACGCCCTCGGACGCGGTGAGTCTCTTGGCGAAGTCGGGGTCCACGACG 387  
 QY 244 GCGTAGTCCACCGCGTCCGGGTGAACGCGTCCGGGTGCGCACCGGGATGTCGTCGG 303  
 Db 386 GAGGCGAGTCTGGGGTGGCTCTGATAGGCGACGCGGATCGGTCAGGAACTTCGCGCGG 327  
 QY 304 GGGGTGAGCGCGCCCTGCTTGGCCGCGTGTGTCGACACCCAGGAGACGAGTCCGGA 363  
 Db 326 ATGATGAGCAGCACCTTCAGATGCGCGCAGCGCCCGAGGTTCGACACTTCATC 267  
 QY 364 CGATCGCGCAGAGTTCGTCAGGTGGCGCTCTTCGCGTCCGCGTACGCCACCCACC 423  
 Db 266 TCACGCTCTTGGCCTTGATGTTTCATGCCCTGCGGTACGAGATCTTGTGGCG 207  
 QY 424 CGCTTGCCTCGGCTTGACGCGAGTTTCAGCAGGCG 458  
 Db 206 TACAGCGCGCGGACGCTGTTGATCAGTGGCG 172

**RESULT 9**  
**AV619788/c**  
**LOCUS** AV619788 532 bp mRNA linear EST 07-NOV-2001  
**DEFINITION** AV619788 Chlamydomonas reinhardtii 5% to 0.04% CO<sub>2</sub> Chlamydomonas reinhardtii cDNA clone LC015e02\_r 5', mRNA sequence.  
**ACCESSION** AV619788  
**VERSION** AV619788.1 GI:10768963  
**KEYWORDS** EST.  
**SOURCE** Chlamydomonas reinhardtii  
**ORGANISM** Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

**REFERENCE**  
**AUTHORS** Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y., and Tabata, S.

TITLE Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii

JOURNAL DNA Res. 7 (5), 305-307 (2000)

MEDLINE 20539644

PUBMED 11089912

COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
source  
Location/Qualifiers  
1..532  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="LC015e02\_r"  
/clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
/note="Vector: pBluescript II SK-; Site: 1: EcoRI; Site 2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

ORIGIN

Query Match 12.0%; Score 59.8; DB 9; Length 532;  
Best Local Similarity 48.7%; Pred. No. 11;  
Matches 163; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 124 CCGCGGGTACGTCGTACCGGGGACGTACAGAGTCCACTGTCGCCACGC 183  
Db 375 CCCAGGAGGAGTATGAGGGATGGCAGCGCTGCGTATGGACAGCGGCCACG 316

QY 184 CGCGGAACCTCTGCTCCATGATCTGTCGGCGTGTCCAGCGGAGAGCAGC 243  
Db 315 CGGCGCCAGGACGGCTCGGACGGTGTGCGAGTCTTGGCGAGTCCAGCAGC 256

QY 244 CGGTAGTCCACCGCTCGGGGTGAACCGCTCCGGGTGCGACCGGATGTCGTCGC 303  
Db 255 GAGGCGACTCGGGTTCGCTGCTAGCGACGCGATGCGGTCCAGMACTGCGCGG 196

QY 304 GGGGTGACCGCGCCCTTGGCGCGGTGTGCGACACCCAGGAGACAGGTCGCGA 363  
Db 195 ATGATGACGACACCTTCAGATGCGCGCCAGCGCGCCAGTCCACTTCATC 136

QY 364 CGGTAGTCCACCGCTCGGGGTGAACCGCTCCGGGTGCGACCGGATGTCGTCGC 423  
Db 135 TCCAGCTCTTGGCTTGTATGTTTCATGCGCTGCGGTACGAGAGATCTGCTGGC 76

QY 424 CGCTTGCCCTCGGCTTGAAGCGAGTTACAGAGGC 458  
Db 75 TACAGCGCGCGGACGCTGCTGATCAGCTGCGC 41

RESULT 10  
BM002630/c  
LOCUS 599 bp mRNA linear EST 25-OCT-2001  
DEFINITION 1031105G06.y1 C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BM002630  
VERSION BM002630.1 GI:16437410  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
REFERENCE Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.  
AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.  
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1031  
JOURNAL Unpublished (2001)

COMMENT

FEATURES  
source  
Location/Qualifiers

1..599  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strains="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

ORIGIN

Query Match 12.0%; Score 59.8; DB 12; Length 599;  
Best Local Similarity 48.7%; Pred. No. 11;  
Matches 163; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 124 CCGCGGGTACGTCGTACCGGGGACGTACAGAGTCCACTGTCGCCACGC 183  
Db 495 CCCAGGAGGAGTATGAGGGATGGCAGCGCTGCGTATGGACAGCGGCCACG 436

QY 184 CGCGGAACCTCTGCTCCATGATCTGTCGGCGTGTCCAGCGGAGAGCAGC 243  
Db 435 CGGCGCCAGGACGGCTCGGACGGTGTGCGAGTCTTGGCGAGTCCAGCAGC 376

QY 244 CGGTAGTCCACCGCTCGGGGTGAACCGCTCCGGGTGCGACCGGATGTCGTCGC 303  
Db 375 GAGGCGACTCGGGTTCGCTGCTAGCGACGCGATGCGGTCCAGMACTGCGCGG 316

QY 304 GGGGTGACCGCGCCCTTGGCGCGGTGTGCGACACCCAGGAGACAGGTCGCGA 363  
Db 315 ATGATGACGACACCTTCAGATGCGCGCCAGCGCGCCAGTCCACTTCATC 256

QY 364 CGGTAGTCCACCGCTCGGGGTGAACCGCTCCGGGTGCGACCGGATGTCGTCGC 423  
Db 255 TCCAGCTCTTGGCTTGTATGTTTCATGCGCTGCGGTACGAGAGATCTGCTGGC 196

QY 424 CGCTTGCCCTCGGCTTGAAGCGAGTTACAGAGGC 458  
Db 195 TACAGCGCGCGGACGCTGCTGATCAGCTGCGC 161

RESULT 11  
BF860172/c  
LOCUS 786 bp mRNA linear EST 19-JAN-2001  
DEFINITION 963015G07.y1 C. reinhardtii CC-1690, Stress condition I, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BF860172  
VERSION BF860172.1 GI:12250309  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonas.

REFERENCE  
1 (bases 1 to 786)  
Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,  
Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants; Project Phase 3  
Unpublished (2000)  
Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.

JOURNAL  
COMMENT

FEATURES  
source  
1..786  
Location/Qualifiers  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Stress condition I,  
normalized, lambda Zap II"  
/notes="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI; This library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,  
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min,  
1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was  
purified from each sample, pooled and cDNA synthesized.  
The cDNA was directionally cloned into lambda Zap II  
(Stratagene) in the EcoRI (5') and XhoI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
Zap clones by superinfection with ExAssist (Stratagene)  
phage. The library was normalized using method 4 described  
in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN  
Query Match 12.0%; Score 59.8; DB 10; Length 786;  
Best Local Similarity 48.7%; Pred. No. 11;  
Matches 163; Conservative 0; Mismatches 172; Indels 0; Gaps 0;  
QY 124 CCGCGCGTCTCTCAGTGTACGCGGGGACGTACAGATCCAGTCTCGGCAGCC 183  
Db 572 CCAGGGAGAGTCTATGAGGGGATGGGACGCCGTGCTGATGACAGCGCGCCAG 513  
QY 184 CGCGGAACTCTCTGCTCTCTGCGCATGATCTGTCGCGGTGCTTCCAGCGAAGAGCAGC 243  
Db 512 CGCGCGCAGGACGCTCGGACGGGTGAGTCTCTTGGCGAAGTCTCGGCTCCACGACGC 453  
QY 244 GGTATCTCACCGCGTGGCGGTGAACGGGTGCGGGGTGCGCACCGGGGATGTGCGTGGCG 303  
Db 452 GAGGGCAGCTCGGGGTGGCGTCTGTAGGGGACCGGATGCGGTCCAGGAACCTGCGCGGG 393  
QY 304 GGGGTGAGCGGCGCTTGTGGCGCGTGTGTCGCACACCCAGAGACAGGTCCGGA 363  
Db 392 ATGATGACGACACCTTTCAGATGCGCGGACGCGCCAGGTGACAGTTCACCTTCATC 333  
QY 364 CGATGCGCAGAGTTCGTACGGTGGCGCTTTCGCGGTCCGCGGTACGCCACAC 423  
Db 332 TCACGCTCTTGGCTTGTATGTTTCATGCGCTGCGGTACGAGCAGATCTTGTGGCG 273  
QY 424 CGCTTGGCTTGGCGGTGAGCGGTTCACAGGGC 458  
Db 272 TACAGCGCGCGGACGTCGTTGATCAGTGGC 238

RESULT 12  
CNS035KP 1085 bp DNA linear GSS 01-SEP-2000  
LOCUS  
DEFINITION  
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
214101 of library G from Tetraodon nigroviridis, genomic survey

sequence.  
AL228850 GI:7887843  
GSS: genome survey sequence.  
KEYWORDS  
Tetraodon nigroviridis  
Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontoidea; Tetraodontidae; Tetraodon.

REFERENCE  
1  
Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,  
Bernot A., Bizet C., Wincker P., Brottier P., Quetier F.,  
Saurin W. and Weissenbach J.  
Estimate of human gene number provided by genome-wide analysis  
using Tetraodon nigroviridis DNA sequence  
Nat. Genet. 25 (2), 235-238 (2000)

JOURNAL  
MEDLINE  
PUBMED  
20296633  
10835645

REFERENCE  
2  
Roest Crolius H., Jaillon O., Dasilva C., Ozouf-Costaz C.,  
Fizames C., Fischer C., Bouneau L., Billault A., Quetier F.,  
Saurin W., Bernot A. and Weissenbach J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Genome Res. 10 (7), 939-949 (2000)

JOURNAL  
MEDLINE  
PUBMED  
20359837  
10899143

REFERENCE  
3 (bases 1 to 1085)  
Direct Submission  
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage  
- Web : www.genoscope.cns.fr  
- BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
Location/Qualifiers  
1..1085  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="214101"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG214AB01SP1-end :  
PUC-Ori"

ORIGIN  
Query Match 11.9%; Score 59.4; DB 29; Length 1085;  
Best Local Similarity 49.5%; Pred. No. 13;  
Matches 153; Conservative 0; Mismatches 156; Indels 0; Gaps 0;  
QY 192 CTCTGTCTCTTCCCATGATCTGTCGGGTGTTCCAGCGAAGAGCAGCGGTAGTC 251  
Db 50 CTCTGTCTCTGCGCTCGCGAAGTAGACAGCGCTGGAGAAGAGATGATCCCGATGAGAG 109  
QY 252 CACCGCTGCGGGGTGAACCGGTCGCGGTGCGCACCGGGATGTGCTGCGGGGTGAG 311  
Db 110 GAAGAAGATGACGAGGCGCCAGCTGCGCATGCTGGCTTCAGGGTCTGGGCCAGGATCTG 169  
QY 312 CCGGCCCTGTGTCGCGCGTGTGTCGCACACCCAGAGACAGGTCGCGGACCGATGCC 371  
Db 170 GAGCCCTTGGAGTGGCGGACAGCTTGAAGTGGGAAGACGCGGACGAGGATGAC 229  
QY 372 GCAGAAATTGTCAGGTGCGGTCTTTCGCGGTGCGGCGGTACGCCACACCGGCTTGGC 431  
Db 230 CTTGAGGATGCGCAGGAGGTGGCTCTGTCGCCACCGCCCTTGGCCCTCTTGTGGC 289  
QY 432 CTGCGGCTTACGAGGTTTACGAGGCGGACAGGTGCGGTGCGGATGCGCTCGCGCTGGC 491  
Db 290 CTGTCGCCAGCTCCCTCCCGAGGTGATGAGTAGGGGATGATGCCACGATGTCAT 349

```

QY 492 GCGAACCT 500
Db 350 GGTGTTAT 358

RESULT 13
CNS00720/c 932 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL066742
VERSION AL066742.1 GI:4945205
KEYWORDS GSS.
ORGANISM Drosophila melanogaster (fruit fly)
SOURCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 932)
AUTHORS Genoscope.
TITLES Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aarton Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             source
    Location/Qualifiers
        1..932
            /organism="Drosophila melanogaster"
            /mol_type="genomic DNA"
            /db_xref="taxon:7227"
            /clone="BACR14B09"
            /clone_lib="RPCI-98"
            /note="end : 17"

ORIGIN
Query Match      11.8%; Score 59; DB 29; Length 932;
Best Local Similarity 29.3%; Pred. No. 15;
Matches 139; Conservative 104; Mismatches 232; Indels 0; Gaps 0;

QY 1 GGTACCCGAGCTGTCCCGAACACGAGTCGAGATACGGCGGAGAGAACACCCCGGT 60
Db 907 GSGCGSSGSGSCCGCCGSSSCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 848

QY 61 AGTCGCGGTAGACGCTGCGCGCGAGAGCGTACGCGCTTCGACGCTCAGCGCGGCGG 120
Db 847 CGCSGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 788

QY 121 ACACGCGGCGGTACGCTGTCAGCTGTCAGCGGCGGAGTCAGATCCACTGTCGCGCA 180
Db 787 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 728

QY 181 GCCCGCGGAGACTCTGCTCTCCATGATCTCGTCGGCGGTGTTCCAGCGGAGAGC 240
Db 727 GSSGCGSGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 668

QY 241 ACGGCGTAGTCCACCGCTCGCGGCGGTGAACGGCTCGGGGTGCGCAGCGGATGTGCTG 300
Db 667 CSSGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 608

```

```

QY 301 CGCGGGGTAGCGCGCCCTGCTTGGCGCGGTGCTGTCACACCCAGGACGAGTCC 360
Db 607 GSGSCCGCGCMVARMVAVSVSCCCSCCMASCCCGCVSGSCSCSCSCSCCGCA 548

QY 361 GGACGATGCCCGCAGAGTGTGTCACGGTGGCGCTCTTCGCGGTGCGCGGTACGCCACC 420
Db 547 SCVTCAGCCNSAGCMCGCMCCCGCCGSCSCSCCGCGCGCGCGCGCGCMVCGCM 488

QY 421 ACCCGCTTCCCTCGCGCTTTCAGCGAGTTCACAGGCGGAGCAGGTCGCTGCGGA 475
Db 487 CCGMGVACCMCGCGVGMCMGCMGRCVSSAGSGCAGMGARMAACAGSGCAGM 433

RESULT 14
BF860173/c 750 bp mRNA linear EST 19-JAN-2001
LOCUS 963015G07.y2 C. reinhardtii CC-1690, Stress condition I,
DEFINITION normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA
sequence.
ACCESSION BF860173
VERSION BF860173.1 GI:12250310
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
REFERENCE 1 (bases 1 to 750)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauser,C., Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLES Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 3
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES             source
    Location/Qualifiers
        1..750
            /organism="Chlamydomonas reinhardtii"
            /mol_type="mRNA"
            /strain="CC-1690 wild type mt+ 21gr"
            /db_xref="taxon:3055"
            /clone_lib="C. reinhardtii CC-1690, Stress condition I, normalized, Lambda Zap II"
            /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN
Query Match      11.8%; Score 58.8; DB 10; Length 750;
Best Local Similarity 48.8%; Pred. No. 16;
Matches 159; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 133 CAGCTCGTCACGTGTACGGCGGAGTACAGATCCACTGTCCGCGACCGCGGGAGAC 192
Db 567 CAGGTATGAGGGGATGGGACACCGTCTGTGTGATGACAGCGCGGCGGCGGCG 508

QY 193 TCCTGCTCTTCGCCCATGATCTGTCGGCGGTGTTCCAGCGGAGAGCGCGTAGTCC 252

```

```

Db      507  GAGCGCTCGGACAGCGGTGAGCTCTTTGGGAACTCGGGGTCCACAGCAGGAGGCGAGC 448
Qy      253  ACCGCGTCGGGCGTGAACGGTCCGGGTGCGCACCGGGATGTGCTGCGGGGTGAGC 312
Db      447  TCGGGGTTCGGCTCGGTAGGCGACGCGGATGCGGTCCAGGAACCTGCGCGGGAATGATGCGAG 388
Qy      313  CGGCGCTGCTTGGCCGGCGTGTGTGTCGCACACCCAGGAGACAGGATCCGAGATGCGG 372
Db      387  CCACCCCTCCAGATGCGGCGCAGCGCCGAGGTCCAGCTTCACATCTCCAGCTC 328
Qy      373  CAGAACTTCGTACAGTGGGCTCTTCGCGTTCGCGCGGTAGCCACACCGCTTGCCC 432
Db      327  TTGGCCTTGATAATGTTCAATGTCCTCGCGGTACGAGCAGATCTTGTGCGGTACAGCGCC 268
Qy      433  TCGGCGCTTACGCGAGTTACGACGGGC 458
Db      267  GCGGCGACGTGCTTGATCAGTGC 242

```

```

RESULT 15
AG043040/c
LOCUS      918 bp      DNA      linear      GSS 01-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-021D18.R, genomic survey sequence.
ACCESSION AG043040
VERSION AG043040.1 GI:16571765
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

```

```

REFERENCE
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(BE-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the R&D process and may have higher chance of
clone-tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
FEATURES
source Location/Qualifiers
1. 918
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-021D18.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match 11.7%; Score 58.6; DB 29; Length 918;
Best Local Similarity 46.2%; Pred. No. 17;
Matches 181; Conservative 0; Mismatches 211; Indels 0; Gaps 0;
Qy 53 CCCCCGCTAGTCGGGTAGACGTGTGGCGGAGAGGCGTACGGCGCTTCGACCGGTACGGG 112
Db 836 CCCCCCGCGCGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777

```

```

Qy      113  CGCGGCGGACACACCGCGGCTCAGCTCGTACGCGGGGACGTACAGGATCCACT 172
Db      776  GCGGCGCGGCGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
Qy      173  GTCCGCGAGCCCGCGGAACTCTGTCTTTCGCGCATGATCTCGTCCGGGTGTTCCAGG 232
Db      716  GGCCNCCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657
Qy      233  CGAAGAGCAGCGCGTAGTTCACACCGCTCGGCGGTGAACGCGTCCGGGTGCGCACCGGGA 292
Db      656  CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597
Qy      293  TGTGCGTCCCGGCGGTGAGCGCGCGCTGTCTTGGCGCGCGTCTGTCGCACACCGAGAGA 352
Db      596  CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 537
Qy      353  CCAGGTCCGAGACCGATGCGCGCAGAGTTCTGTCAGGTGGCGCTCTTCGCGCGTCCGCGGT 412
Db      536  GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 477
Qy      413  ACGCCACACCGCGCTTCCCTCGCGCTTGAGC 444
Db      476  CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 445

```

```

RESULT 16
CK207202/c
LOCUS      1101 bp      mRNA      linear      EST 08-DEC-2003
DEFINITION FGAS018821 Triticum aestivum FGAS: Library 5 GATE 7 Triticum
aestivum cDNA, mRNA sequence.
ACCESSION CK207202
VERSION CK207202.1 GI:39569592
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
AUTHORS Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilson, D.,
Penniket, C., Roach, J.L. and Sarhan, F.
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL Unpublished (2003)
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas_est@cs.usask.ca
This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [39,775].
Plate: LSB010 row: A column: 10.
FEATURES
source Location/Qualifiers
1. 1101
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone_lib="Triticum aestivum FGAS: Library 5 GATE 7"
/note="Vector: pCMV.SPORT6; Crown and developmental stages
of spike formation in wheat cultivar Norstar 4 mRNA
populations were combined before constructing the library.
The first mRNA population is from 1cm crown sections after
30 days of cold acclimation. The second is from 1cm crown
sections after 11 days of deacclimation (before
deacclimation plants were fully vernalized for 49 days).
The third is from different developmental stages of spike

```



```

/clone="BACR14B09"
/clone_lib="RPCI-98"
/notes="end : T7"

ORIGIN
Query Match      11.6%; Score 58; DB 29; Length 932;
Best Local Similarity 32.6%; Pred. No. 21;
Matches 127; Conservative 84; Mismatches 178; Indels 1; Gaps 1;

QY 57 CGGTAGTCGGGTAGACGCTGGCGGAGAGGGGTACGCGCTTCACGGTACAGCGGGG 116
Db 542 CTGBGCTGCGGGGSGSGGSGCGBGCGGGGSGTGGSGGGGSGSBTKYTBKRGKGGG 601
QY 117 GCGGACACCGCGGGGTACGCTGCTCAGCTGTACGCGGGGAGGTACAGGATCACTGTCC 176
Db 602 SSGSCCGSGSGCGGGGCGSGSGSGCGGSGSGCGSCGSGSGGCGCGCGCGSGS 661
QY 177 GCGAGCGCGCGGAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 235
Db 662 GCGSGGSGSGSGCGCGGSGSGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 721
QY 236 AGACGACGCGGTAGTCCACCGGTGCGGGGCGTGAACGCGTCCGGGTGCGCACCGGATGT 295
Db 722 GCGSCGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
QY 296 GCGTCCCGGGGTGACCGCGCGCTGCTTGGCGCGCGCTGCTGCGCACACCGAGACCA 355
Db 782 CGCGCGSGSGSGCGSGSGSGCGSGSGSGCGSGSGSGCGSGSGSGSGSGSGSG 841
QY 356 GGTCCGCGACCGATGCGCGAGAGTTCGTACCGGTGCGCTCTTCCGCGCGCGGTAG 415
Db 842 CSBGCGCGCGSGSGCGSGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSG 901
QY 416 CCACCGACCGCTTGCCTCGCGCTTGAAGCG 445
Db 902 SGGCGCGCGSGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 931

RESULT 19
CNS006XX
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
fly); genomic survey sequence.
AL066051
GI:4945019
GSS.
Drosophila melanogaster (fruit fly)
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 935)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Oosagawa and
Aaron Marmore in Pierer de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES
source
Location/Qualifiers
1..935
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR14N09"
/clone_lib="RPCI-98"
/notes="end : T7"

ORIGIN
Query Match      11.6%; Score 58; DB 29; Length 935;
Best Local Similarity 28.5%; Pred. No. 21;
Matches 109; Conservative 105; Mismatches 167; Indels 1; Gaps 1;

QY 115 GGGCGGACACCGCGGGTCACTGCTCAGCTGTACGCGGGGAGGTACAGGATCCACTGT 174
Db 495 VSSGTGTCMDMVMVCGCGSCMCRTRSSGSGTCTCMYMSVSVSCSGTCCGTCSSCTSC 554
QY 175 CCGCAGCCCGCGGGAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 234
Db 555 KMCSTCTGCGCGCGGCTSCSSSSCCBSBSYTCSTCTCTKSSGCTSGGCTGCCGG 614
QY 235 AAGAGCAGCGGTAGTCCACCGGTGCGGGTGAACGCGTCCGGGTGCGCACCGGATG 294
Db 615 GSGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 674
QY 295 TGCGTGC CGGGGTGAGCGCGCGCTGCTTGGCGCGCGCTGCTGCGCACACCGAGGACC 354
Db 675 GSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 734
QY 355 AGGTCCGACCGTCCCGCGAGAGTTCGTACCGGTGCGCTCTTCCGCGTCCGCGCTAC 414
Db 735 GCSCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 794
QY 415 CCACCGACCGCTTGCCTCGCGCTTGAAGCGTTCAGCAGCGGCGAGC-AGGTGCGTGC 473
Db 795 CSBGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 854
QY 474 GATCCCTTCGACGTCGCGCGCG 495
Db 855 SGGSCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 876

RESULT 20
BZ365371
LOCUS
DEFINITION
ie1se07.g2 WGS-ZmaysF (JMI07 adapted methyl filtered) Zea mays
genomic clone ie1se07 5', genomic survey sequence.
BZ365371
GI:25074825
GSS.
Zea mays
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 588)
Rabinowicz, P.D., O'Shaughnessy, A.L., Baliya, V., Dedhia, N.,
Karzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
Zucavern, J., McCombie, W.R. and Martienssen, R.A.
Genomic shotgun sequences from Zea mays (methyl-filtered)
Unpublished (2002)
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: ie15 row: e column: 07
Seq primer: -2M13UnivRev
Class: shotgun
High quality sequence stop: 588.

```

FEATURES  
source

Location/Qualifiers  
1. .588  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ze15e07"  
/lab\_host="JM107 or DH5a"  
/note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;  
The vector was digested with XbaI and one nucleotide was  
added by fill in in the recessive 3' end. The genomic DNA  
was nebulized, end repaired, adaptor ligated and size  
fractionated using sephadex. The resulting fragments were  
between 0.8 and 3 kb and were cloned into the vector  
(.x/y reads in M13mp19, .b/g reads in pUC19). The same  
ligation was transformed in either JM107 or DH5a. "

## ORIGIN

Query Match 11.4%; Score 57; DB 28; Length 588;  
Best Local Similarity 50.5%; Pred. No. 29;  
Matches 138; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 188 GGAATCTCTGCTTCCTCCATGATCTGTCGGCGTGTTCACGGCGAAGACGACGCGT 247  
DB 131 GGTCCACCACTCCCGGTCGGACCTCTCGCGTCTCTCCACCTTGATGATCTCGT 190

QY 248 AGTCACCGCTCGGCGTGAACGCTCCGGGTGCGCACCGGATGTGTCGCGGGG 307  
DB 191 AGTAGCATGTTGGGAGCAGCGGTAGGCGACGTCTCGCGCGTTCAGCGGGGGA 250

QY 308 TGAGCGGCGCTGCTTGCGCGGCTGCTGTGCGCACCCAGGAGACCAAGTTCGCGACCGA 367  
DB 251 GGTGTAGTCCGAGTAGCACTCCGAGGAGGCGTACATGTTGACACCGGCGGCGC 310

QY 368 TGCGCAGAGTTCGTACGCTGCGGCTCTTCGCGGTGCGCGGTACGCCACCGCGT 427  
DB 311 CGCGTAGACTCCAGCGGGCGACGTACTCGCGCATGGACCCCGTACGATGACGTCGA 370

QY 428 TGCCCTCGGCTTGAGCGGAGTTCAGCAGGGCGA 460  
DB 371 TGTACTTGTGTCGGCGCCAGAGCGCGCACGA 403

## RESULT 21

CC334644  
LOCUS  
DEFINITION  
OGAM79TH ZM 0.7.1.5 KB Zea mays genomic clone ZMMBma0367M13,  
genomic survey sequence.

ACCESSION  
CC334644.1 GI:30804057

VERSION  
GSS.

KEYWORDS  
Zea mays

SOURCE  
Zea mays

ORGANISM  
Zea mays

REFERENCE  
AUTHORS  
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE  
JOURNAL  
COMMENT  
Unpublished (2002)

CONTACT: Cathy Whitelaw

7912 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .755

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBma0367M13"

/note="Vector: pBCK-; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBma0367M13"  
/clone\_lib="ZM 0.7.1.5 KB"  
/note="Vector: pBCK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 11.4%; Score 57; DB 28; Length 755;  
Best Local Similarity 50.5%; Pred. No. 29;  
Matches 138; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 188 GGAATCTCTGCTTCCTCCATGATCTGTCGGCGTGTTCACGGCGAAGACGACGCGT 247  
DB 336 GGTCCACCACTCCCGGTCGGACCTCTCGCGTCTCTCCACCTTGATGATGAACTCGT 395

QY 248 AGTCACCGCTCGGCGTGAACGCTCCGGGTGCGCACCGGATGTGTCGCGGGG 307  
DB 396 AGTAGCATGTTGGGAGCAGCGGTAGGCGACGTCTCGCGCGTTCAGCGGGGGA 455

QY 308 TGAGCGGCGCTGCTTGCGCGGCGTGTGTCGACACCCAGGAGACCAAGTTCGCGACCGA 367  
DB 456 GGTGTAGTCCGAGTAGCACTCCGAGGAGGCGTACATGTTGACACCGGCGGCGC 515

QY 368 TGCGCAGAGTTCGTACGCTGCGGCTCTTCGCGGTGCGCGGTACGCCACCGCGT 427  
DB 516 CGCGTAGACTCCAGCGGGCGACGTACTCGCGCATGGACCCCGTACGATGACGTCGA 575

QY 428 TGCCCTCGGCTTGAGCGGAGTTCAGCAGGGCGA 460  
DB 576 TGTACTTGTGTCGGCGCCAGAGCGCGCACGA 608

## RESULT 22

CG275728  
LOCUS  
DEFINITION  
OGXFI14TH ZM 0.7.1.5 KB Zea mays genomic clone ZMMBma0620D03,  
genomic survey sequence.

ACCESSION  
CG275728.1 GI:34187869

VERSION  
GSS.

KEYWORDS  
Zea mays

SOURCE  
Zea mays

ORGANISM  
Zea mays

REFERENCE  
AUTHORS  
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE  
JOURNAL  
COMMENT  
Unpublished (2002)

CONTACT: Cathy Whitelaw

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .772

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBma0620D03"

/clone\_lib="ZM 0.7.1.5 KB"

/note="Vector: pBCK-; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"



[illegible][illegible]

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [Strasberg-remail@Cgap.Stanford](mailto:Strasberg-remail@Cgap.Stanford)  
The  
DNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

```

FEATURES
    source
        1..948
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6298042"
            /tissue_type="hepatocellular carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_100"
            /notes="Organ: liver; Vector: pOTF7; Site 1: XhoI; Site 2:
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Cloning Strategy: Lambda ZAP Express
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2508 row: h column: 11
High quality sequence stop: 398.
Location/Qualifiers
```

```

/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:

```

EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

```

ORIGIN
Query Match      11.4%; Score 56.8; DB 13; Length 948;
Best Local Similarity 46.1%; Pred. No. 32;
Matches 175; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 64 CCGGCTAGACCGTGGCGCGGAAGCGGTACGCGCTTCAGCGGTACGCGCGCGCGGACACA 123
DB 546 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 605
QY 124 CCGGCGCGGTACGCTGTCATCGTACCGCGGCGGACGTACAGATCCATCTCTCCGACGCC 193
DB 606 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 665
QY 184 CCGGCGGAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 243
DB 566 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 725
QY 244 CCGTAGTCCACCGGTGCGCGGTGAACCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTG 303
DB 726 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 785
QY 304 GGGGTGACCGCGCGCTTTGGCGCGGTGTCGCGCACACCGGAGACACAGTCCGGA 363
DB 786 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 845
QY 364 CCGTAGTCCACGAGATTCGTCAACGCGTGGCGCTCTTCGCGCGGTGCGCGGTGCGCGGT 423
DB 846 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 905
QY 424 CGCTGCGCTCGGCTTCAG 443
DB 906 GNCGGCGNCCGCGCGCG 925

RESULT 25
AL820302/c
LOCUS AL820302 N:130 Triticum aestivum cDNA clone H01.N130.Plate_48, mRNA
DEFINITION sequence.
ACCESSION AL820302
VERSION AL820302.1 GI:21831902
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 503)
AUTHORS Wilson,I., Bewick,R., Shepherd,S., Barker,G., Parker,J., Owen,P.,
Edwards,D., Coghill,J., Holdsworth,M., Lenton,J., Shewry,P. and
Edwards,K.
A BBSRC-funded wheat EST resource for the academic community
Contact: Barker G
Institute of Arable Crop Research
Long Ashton, Bristol BS41 9AF United Kingdom.
Location/Qualifiers
1. 503
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="mercia"
/db_xref="taxon:4565"
/clone="H01.N130.Plate_48"
/tissue_type="embryo"

FEATURES
Source
1. 503
/organism="Triticum aestivum"
/mol_type="mRNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0526F13"
/clone_lib="ZM.0.7.1.5_KB"
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      11.2%; Score 55.8; DB 29; Length 779;
Best Local Similarity 48.7%; Pred. No. 45;
Matches 213; Conservative 0; Mismatches 217; Indels 7; Gaps 2;

/dev stage="2 days post germination"
/clone_lib="N:130"

ORIGIN
Query Match      11.2%; Score 55.8; DB 9; Length 503;
Best Local Similarity 51.4%; Pred. No. 43;
Matches 129; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 249 GTCCACCGCGTGGCGGTGAACGGTTCGGGGTGGCACCGGGATGTCGTGCGCGGGGT 308
DB 289 GTTCGCGCATTTGGTATGAACAATCTGGGAATCCATCCGGAATCTCTCCAGGAGAT 230
QY 309 GAGCGCGCGTGGCGCGGTGTCGTCGACACCGGAGAGACCGAGTCCGAGACCGAT 368
DB 229 GGTGCAGAGCTCGATGTCGTCTCCGGCGCGCCACCTGCTGGCGCGAGATCCCTCATCTC 170
QY 369 GCGCGAGAAGTTGTCACGCTGGCGCTCTTCGCGTGGCGCGGTACGCCACACCGCGTT 428
DB 169 CTCGAGTGTAGACGAGATGCTCTCTGTCGCCGCGCGCGCGCGCGCATCTTGGGAC 110
QY 429 GCGCTCGCGCTTGACGAGTTTCAGCAGGCGGAGAGGTCGTCGCGGATGCGCTCGACGTC 488
DB 109 GCGCAGCGCTTACGTCGGGTGTCAGCACGTGGAGCGCGTGGCGGTGGCGGTGCCGAC 50
QY 489 GCGCGCGGACCC 499
DB 49 GCGCGGAGCCC 39

RESULT 26
CG324766
LOCUS CG324766
DEFINITION genomic survey sequence.
ACCESSION CG324766
VERSION CG324766.1 GI:34242032
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 779)
AUTHORS Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
COMMENT Other_GSSs: OGWBE31TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
Location/Qualifiers
1. 779
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0526F13"
/clone_lib="ZM.0.7.1.5_KB"
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      11.2%; Score 55.8; DB 29; Length 779;
Best Local Similarity 48.7%; Pred. No. 45;
Matches 213; Conservative 0; Mismatches 217; Indels 7; Gaps 2;

```





/clone\_lib="PTB Chimpanzee Male BAC Library"

## ORIGIN

Query Match 11.1%; Score 55.4; DB 29; Length 986;  
Best Local Similarity 44.7%; Pred. No. 53;  
Matches 200; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

QY 53 CCCCGGTATGTCGGGTAGACGGTGGCGGAGGCGGTACGGCCCTTCAGACGTCAGCGG 112  
|||  
Db 437 CGCGCGGGGCGCGGGGCGCGGCCCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496  
|||

QY 113 GCGGGCGGACACCGGCGGGTACGTGTACGTGTACGGGGGAGGTACAGATCCACT 172  
|||  
Db 497 GCGCGCGGGCG 556  
|||

QY 173 GTCCGCGAGCCCGCGCGAACTCTCTCTCTTCCATGATCTCTCGGCGGTGTTCAGG 232  
|||  
Db 557 GCCCGCCCG 616  
|||

QY 233 CGAAGACAGCGGTAGTTCACCGCGTGGGGGTGAACCGGTTCGGGGTTCGACCGGGA 292  
|||  
Db 617 CGCGGGGCG 676  
|||

QY 293 TGTGCGTGC CGGGGTGAGCGCGCCCTCTGTGCGCGGTCTGTGCGACACCGAGGAGA 352  
|||  
Db 677 CGCGGGGCGCGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 736  
|||

QY 353 CAGGTTCGGAACGATGCGCGAGAAATTGTGTACGGTGGCGTCTTCGCGCGCGCGGT 412  
|||  
Db 737 CGCGGGGCG 796  
|||

QY 413 AGCCACACCGCGCTTTCCTCGCGCTTTCGCGAGTTTCAGCGGCGGACAGTCTGGTGC 472  
|||  
Db 797 CG 856  
|||

QY 473 GATGCGCTTCGACGTTCGCGCGGCGAAC 499  
|||  
Db 857 GCG 883  
|||

## RESULT 31

BQ080807/c  
LOCUS  
DEFINITION  
103007C04.y3 C. reinhardtii CC-1690, Deflagellation (normalized),  
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BQ080807  
VERSION BQ080807.1 GI:22049444

KEYWORDS  
SOURCE Chlamydomonas reinhardtii

ORGANISM Chlamydomonas reinhardtii

REFERENCE 1 (bases 1 to 456)

AUTHORS Groisman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C.,

Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants. Project: 1030

JOURNAL Unpublished (2002)

COMMENT Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

1..456

/organism="Chlamydomonas reinhardtii"

/mol\_type="mRNA"

/strain="CC-1690 wild type mt+ 21gr"

/db\_xref="taxon:3055"

/clone\_lib="C. reinhardtii CC-1690, Deflagellation"

## FEATURES

source

(normalized), Lambda Zap II"  
/note=vector; pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI; Deflagellation library, constructed by John Davies  
and Jeffrey McDermott, combines cDNAs from CC-1690 cells  
which had been re-synthesizing flagella for 15, 30 and 60  
min after being deflagellated by pH shock. PolyA mRNA was  
purified from each sample, pooled and cDNA synthesized.  
The cDNA was directionally cloned into lambda Zap II  
(Stratagene) in the EcoRI (5') and XhoI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
ZAP clones by superinfection with ExAssist (Stratagene)  
phage. The library was normalized using method 4 described  
in Bonaldo et al., (1996) Genome Research 6: 791-806."

## ORIGIN

Query Match 11.0%; Score 55.2; DB 13; Length 456;  
Best Local Similarity 53.2%; Pred. No. 52;  
Matches 117; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 124 CCGCGGGTCACTCGTCAGTCGTCACGTGACGCGGGGACGTACAGGATCCACTGTCCGCCAGCC 183  
|||  
Db 280 CCCAGGGAGGAGGTCTATGAGGGGATGGGACCGCGTGCCTGATGGACAGCGCGGCACG 221  
|||

QY 184 CGCGGGAACCT 243  
|||  
Db 220 CGCGGCCAGGACGCTCGGACGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 261  
|||

QY 244 GCCTAGTTCACCGCGTTCGCGGTGAACCGTTCGCGGTTCGCGACCGCGGATGTGCGTCCG 303  
|||  
Db 160 GAGGCGAGCTCGGGTTCGCGTTCGCGACCGCGATCGCGTTCGCGAACTCGCGCGCG 101  
|||

QY 304 GGGGTGAGCGCGCGCT 343  
|||  
Db 100 ATGATGACGACCGACCT 61  
|||

## RESULT 32

BQ080807/c

LOCUS

DEFINITION

103007C04.y3 C. reinhardtii CC-1690, Deflagellation (normalized),

clone HVSMEB0007N02f, mRNA sequence.

ACCESSION BQ080807

VERSION BQ080807.2 GI:13090587

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Wing RA

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Total hg bases = 397

Seq primer: AATTAAACCTCACTAAAGG

High quality sequence stop: 497.

Location/Qualifiers

1..500

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

.....

RESULT 33  
 BX425797/c  
 LOCUS  
 DEFINITION  
 BX425797 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone  
 CLOBB030ZE01 3-PRIME, mRNA sequence.  
 ACCESSION  
 BX425797  
 VERSION  
 BX425797.1 GI:30784485  
 KEYWORDS  
 EST.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

FEATURES	Location/Qualifiers	Source
1..828		
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone="CLOBB0302501"	
	/tissue_type="NEUROBLASTOMA"	
	/clone_lib="Homo sapiens NEUROBLASTOMA"	
	/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."	

ORIGIN

RESULT 34	CB678003/c	CB678003	736 bp	mRNA	linear	EST 09-APR-2003
LOCUS		OSJNBe15106.f	OSJN06	<i>Oryza sativa</i>	(japonica cultivar-group)	cdNA
DEFINITION		clone OSJNBe15106 5'		mRNA sequence.		
ACCESSION		CB678003				
VERSION		CB678003.1	GI:29681728			
KEYWORDS		EST.				
SOURCE		<i>Oryza sativa</i>	(japonica cultivar-group)			
ORGANISM		<i>Oryza sativa</i>	(japonica cultivar-group)			
		Eukaryota;	Viridiplantae;	Streptophyta;	Embryophyta;	Tracheophyta;
		Spermatophyta;	Magnoliophyta;	Liliopsida;	Poales;	Poaceae;
		Ehrhartoideae;	Oryzaceae;	Oryza.		

REFERENCE 1 (bases 1 to 736)  
 AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
 TITLE Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR Primers  
 FORWARD: gta aaa cga cgg cca gta g  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 15 row: I column: 06  
 Seq primer: gta aaa cga cgg cca gta g.  
 FEATURES  
 source  
 1. 736  
 Location/Qualifiers  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNE16N22"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEe"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)"  
 ORIGIN  
 Query Match 11.0%; Score 55; DB 14; Length 736;  
 Best Local Similarity 50.2%; Pred. No. 59;  
 Matches 136; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 QY 189 GAACCTCTGCTCCCTTCGCCATGATCTCGTCGGCGTGTTCACGCGAAGACGAGCGCGTA 248  
 Db 357 GAACCCCGGTTCTTGAGATGTTCCCTTGGCCAGCGTCGGCGGTACAGCCCTGAT 298  
 QY 249 GTCCACCGCGTCGGCGGTGAACCGCTCCGGGTGCGCACCGGATGTGCTGCCGGGGT 308  
 Db 297 GGCACGCGCTCGATGATGACGGCGCGCACGCGGGTCTCTCCACGCCCGGGTTGGAT 238  
 QY 309 GAGCGCGCTGCTTGGCGCGGTCGTGCGCACCCAGGACGAGTCCGGACCGAT 368  
 Db 237 GACAGCGCACCTCTGCTCCATCTTCCGCTTGAACGCCCCACGCTACGATCCACCGTT 178  
 QY 369 GCCGACAGAGTTCTACGCTGGCGCTCTTGGCGTGGCGGTACGCCACCCCGCTT 428  
 Db 177 GCTCCGCTAGATGCTCTGATGGGAGCACCCCACTCGGGGCTCACCACACGTTGAG 118  
 QY 429 GCCCTCGGCTTGAGCGAGTTCAGCAGGGCG 459  
 Db 117 CTGCTCCGCTTGGCGCAGGTGCGCGCGG 87  
 RESULT 35  
 CB651766/c  
 LOCUS 774 bp mRNA linear EST 08-APR-2003  
 DEFINITION OSJNE16N22.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA  
 clone OSJNE16N22 5', mRNA sequence.  
 ACCESSION CB651766  
 VERSION CB651766.1 GI:29646759  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 774)

AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
 TITLE Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR Primers  
 FORWARD: gta aaa cga cgg cca gta g  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 16 row: N column: 22  
 Seq primer: gta aaa cga cgg cca gta g.  
 FEATURES  
 source  
 1. 774  
 Location/Qualifiers  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNE16N22"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEb"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (Che 86061)"  
 ORIGIN  
 Query Match 11.0%; Score 55; DB 14; Length 774;  
 Best Local Similarity 50.2%; Pred. No. 59;  
 Matches 136; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 QY 189 GAACCTCTGCTCCCTTCGCCATGATCTCGTCGGCGTGTTCACGCGAAGACGAGCGCGTA 248  
 Db 701 GAACCCCGGTTCTTGAGATGTTCCCTTGGCCAGCGTCGGCGGTACAGCCCTGAT 642  
 QY 249 GTCCACCGCGTCGGCGGTGAACCGCTCCGGGTGCGCACCGGATGTGCTGCCGGGGT 308  
 Db 641 GGCACGCGCTCGATGATGACGGCGCGCACGCGGGTCTCTCCACGCCCGGGTTGGAT 582  
 QY 309 GAGCGCGCTGCTTGGCGCGGTCGTGCGCACCCAGGACGAGTCCGGACCGAT 368  
 Db 581 GACGAGCGCACCTCTGCTCCATCTTCCGCTTGAACGCCCCACGCTACGATCCACCGTT 522  
 QY 369 GCCGACAGAGTTCTGCTACGCTGGCGCTCTTCCGCTGCGCGGTACGCCACCCCGTT 428  
 Db 521 GCTCCGCTAGATGCTCTGATGGGAGCACCCCACTCGGGGCTCACCACACGTTGAG 462  
 QY 429 GCCCTCGGCTTGAGCGAGTTCAGCAGGGCG 459  
 Db 461 CTGCTCCGCTTGGCGCAGGTGCGCGCGG 431  
 RESULT 36  
 CB672263/c  
 LOCUS 777 bp mRNA linear EST 09-APR-2003  
 DEFINITION OSJNE06D11.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA  
 clone OSJNE06D11 5', mRNA sequence.  
 ACCESSION CB672263  
 VERSION CB672263.1 GI:29675988  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 777)

**AUTHORS** Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E., Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.  
**TITLE** Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea  
**JOURNAL** Unpublished (2003)  
**COMMENT** Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR Primers  
 FORWARD: gta aaa cga cgg cca gtc  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 06 row: D column: 11  
 Seq primer: gta aaa cga cgg cca gtc.

**FEATURES**  
 source  
 1..777  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNE06d11"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEa"  
 /note="Vector: pBluescript II KS +; Site\_1: EcoRI; Site\_2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

**ORIGIN**  
 Query Match 11.0%; Score 55; DB 14; Length 777;  
 Best Local Similarity 50.2%; Pred. No. 59;  
 Matches 136; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 QY 189 GAATCTCTCTCTTGGCCATGATCTCGTCGGCGTGTTCAGCGAAGACGCGGTA 248  
 Db 703 GAACCGCGCTTCTTGGCATGTTCCCTTGGCAGCGTTCGGCGGTACAGCGCCTGAT 644  
 QY 249 GTCCACCGCGTGGCGGTGAACGGTTCGGGTGCGCACCGGATGTGCGTCCGGGGT 308  
 Db 643 GCGGACGCGGTGATAGCGGGCGCCACCGGGTCTCTCCACGCGGGTGTGTGAT 584  
 QY 309 GAGCGCGCCCTCTTGGCGCGGTGCTGTGCACACCCAGGACCGAGTCCGACCGAT 368  
 Db 583 GACGAGCGCCACCTCGTCCATCTTCGCTTGAACGCCACCGGTACGAGTCCACCGCTT 524  
 QY 369 GCGGACGAGTCTGTCACGGTGGCGCTTTCGCGTCCGCGTACGCCACCCCGCTT 428  
 Db 523 GCTCGCGTAGATGGTCTGCATGGGAGCACCCCGCCTCGGGGTCTACCGACACGTTGAG 464  
 QY 429 GCGCTCGGCGCTTACGCGAGTTACGAGGGCG 459  
 Db 463 CTGCTCGCGCTGGCGCAGGTGCGCGCGCG 433

**RESULT 37**  
 CB671209/c  
 LOCUS CB671209 810 bp mRNA linear EST 09-APR-2003  
 DEFINITION OSJNE04K19.f OSJNE Oryza sativa (japonica cultivar-group) cDNA clone OSJNE04K19, mRNA sequence.  
 ACCESSION CB671209  
 VERSION CB671209.1 GI:29674934  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.  
 REFERENCE 1 (bases 1 to 810)  
 AUTHORS Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,

Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.  
 Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea  
 Unpublished (2003)  
 Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR Primers  
 FORWARD: gta aaa cga cgg cca gtc  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 04 row: K column: 19  
 Seq primer: gta aaa cga cgg cca gtc.

**FEATURES**  
 source  
 1..810  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNE04K19"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEa"  
 /note="Vector: pBluescript II KS +; Site\_1: EcoRI; Site\_2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

**ORIGIN**  
 Query Match 11.0%; Score 55; DB 14; Length 810;  
 Best Local Similarity 50.2%; Pred. No. 59;  
 Matches 136; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 QY 189 GAATCTCTCTCTTGGCCATGATCTCGTCGGCGTGTTCAGCGAAGACGCGGTA 248  
 Db 690 GAACCGCGCTTCTTGGCATGTTCCCTTGGCAGCGTTCGGCGGTACAGCGCCTGAT 631  
 QY 249 GTCCACCGCGTGGCGGTGAACGGTTCGGGTGCGCACCGGATGTGCGTCCGGGGT 308  
 Db 630 GCGGACGCGGTGATAGCGGGCGCCACCGGGTCTCTCCACGCGGGTGTGTGAT 571  
 QY 309 GAGCGCGCCCTCTTGGCGCGGTGCTGTGCACACCCAGGACCGAGTCCGAGACCGAT 368  
 Db 570 GACGAGCGCCACCTCGTCCATCTTCGCTTGAACGCCACCGGTACGAGTCCACCGCTT 511  
 QY 369 GCGGACGAGTCTGTCACGGTGGCGCTTTCGCGTCCGCGTACGCCACCCCGCTT 428  
 Db 510 GCTCGCGTAGATGGTCTGCATGGGAGCACCCCGCCTCGGGGTCTACCGACACGTTGAG 451  
 QY 429 GCGCTCGGCGCTTACGCGAGTTACGAGGGCG 459  
 Db 450 CTGCTCGCGCTGGCGCAGGTGCGCGCGCG 420

**RESULT 38**  
 BX442207/c  
 LOCUS BX442207 924 bp mRNA linear EST 15-MAY-2003  
 DEFINITION BX442207 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF02YD08 5-PRIME, mRNA sequence.  
 ACCESSION BX442207  
 VERSION BX442207.1 GI:30786042  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 924)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)



## COMMENT

Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DF027B04QPI.

## FEATURES

source  
Location/Qualifiers  
1. .924  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF027YD08"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

## ORIGIN

Query Match 11.0%; Score 54.8; DB 13; Length 924;  
Best Local Similarity 37.5%; Pred. No. 65;  
Matches 152; Conservative 53; Mismatches 195; Indels 5; Gaps 1;

QY 38 CGGCGAGAGAACACCCCGGTAGTCCGGGTAGAGCGGTGGCGGCGAAGCGGTACCGCC 97  
Db 874 CGGCGGGGNCNCCCGCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 815  
QY 98 TTGACCGGTGAGCGGGCGGCGGACACCGCGCGGTGAGTCTGTCAGTGTACCGGGGA 157  
Db 814 GSGGGGGGGGSCCG 755  
QY 158 GTTACAGATTCAGTTC-----GCCAGCCCGCGGAACTCTGCTCTTCTGCGCATGAT 212  
Db 754 CGSSCGGGGCCSSSSSSCCSSSSGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 695  
QY 213 CTGTCGCGGTGTTCCAGCGCAAGAGACGCGGTGATCCACCGCGTCCGGGTGTAACGC 272  
Db 694 GSGCGCGCGSSCGGCG 635  
QY 273 GTCCGGGTGCGCACCGGATGTCGTGCGGGGTGAGCCGCCCTGTTGCGCGGCT 332  
Db 634 SCGGGCGCGGCGSCCG 575  
QY 333 CGTGTGCGCACACCGAGACAGGTCCGACCGATGCCGCGAGAGTTGTCACGTGCG 392  
Db 574 GGGGGGGVCCCGGCGSSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 515  
QY 393 GCTCTTTCGCGTCCGCGGTACGCCACCGCGTTCGCGTCCGCGC 437  
Db 514 CCGSCG 470

## RESULT 39

AL538546  
LOCUS  
DEFINITION AL538546 Homo sapiens FETAL BRAIN Homo sapiens linear EST 31-MAY-2003  
CS0DF024YJ18 3-PRIME, mRNA sequence.

## ACCESSION

AL538546

## VERSION

1

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

On Feb 13, 2001 this sequence version replaced gi:12802039.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 399.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DF024DE09NP1&cluster=399.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DF024DE09NP1.

## FEATURES

source

Location/Qualifiers  
1. .1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF024YJ18"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

## ORIGIN

Query Match 11.0%; Score 54.8; DB 9; Length 1201;  
Best Local Similarity 32.2%; Pred. No. 67;  
Matches 112; Conservative 81; Mismatches 151; Indels 4; Gaps 1;

QY 152 GGGGAGCGTACAGATCCACTGTCCGCGAGCGCGCGGAACTCTGCTTCCGCCATGA 211  
Db 788 GGGGAGAGGAAGAAGMGGMGGGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMM 847  
QY 212 TCTCGTCGCGTGTTCACGCGAAGAGCAGCGGTAGTCCACCGCGTCCGCGCGTGAACG 271  
Db 848 GMMMMGGGGGMMMMMMGGGCGMMGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 907  
QY 272 CGTCCGGGTGCGCACCGGATGTCGTGCGGGGGGTGAGCCGCCCTGCTTGCCTCCGCG 331  
Db 908 GGGMMGGGGGMMMMGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 967  
QY 332 TCGTGTGCGCACACCGAGACACAGTCCGACCGATCCGCGAGAGTTGCTACCGTGG 391  
Db 968 KCCGCGKCCCCCGCGKCSMMMKSKKMMKKKMMKKKMMKKKMMKKKMMKKKMMKK 1027  
QY 392 CGCTTTCGCGTCCGCGCGTACCGCACCGCGCTTCCCTCGGCTTGCAGCGAGTTCA 451  
Db 1028 SGC-----CCCCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGSG 1083  
QY 452 CGAGGGCGAGCAGTCCGTGCGGATGCCCTCGAGCTCGCGCGCGGAACC 499  
Db 1084 GGGGGGGGCG 1131

## RESULT 40

CA075606/c

LOCUS

DEFINITION

CA075606

5', mRNA sequence.

ACCESSION

CA075606

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

Saccharum officinarum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

clade; Panicoideae; Andropogoneae; Saccharum.

REFERENCE

1 (bases 1 to 605)

AUTHORS

Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.



```

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..905
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM0626G15"
/clone_lib="ZM 0.6 1.0 KB"
/notes="Vector: pCR4-TOPO; Site:1: EcoRI; 0.6-1.0 kb high
CoT selected genomic DNA library"

FEATURES
source
1..905
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM0626G15"
/clone_lib="ZM 0.6 1.0 KB"
/notes="Vector: pCR4-TOPO; Site:1: EcoRI; 0.6-1.0 kb high
CoT selected genomic DNA library"

ORIGIN
Query Match 10.8%; Score 54.2; DB 28; Length 905;
Best Local Similarity 48.8%; Pred. No. 79;
Matches 146; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 199 TCCTTCGCCATGATCGTCGGCGTGTCCAGGCGAAGAGCAGCGAGTCCACCGG 258
Db 439 TCGGCTCTTACAGCGCAGCGCGGCCCTCGGCCCTTACCTTGAGCAGCAG 498
QY 259 TCGGGCGTGAACCGCTCGCGGTTGCGACCGGATGTGCGCGGGTGTGAGCCGCCC 318
Db 499 TCCGCTTGGCACCCTTGTGCGTGTGAGCGGCTTGACGGCGGCGCTTCCGGTCC 558
QY 319 TGCTTGGCGGCTGTGTCGACACCCAGGAGCAGGTCCGACCGATGCGGAGAG 378
Db 559 ACGGGGAACCTTGAAGACCCCGGACCATGAGTGCACCGCGCGTTCAGGGCCCTGGACG 618
QY 379 TTCGTACCGTGGCGCTTTCGCCGTGCGCGGTACGCCACACCGCTTGCCCTCGGCC 438
Db 619 TACACCATGACGGCGTTCGCCGTTCGCTCAGCGGAGCGCGGCCACCTCGTAC 678
QY 439 TTGAGCGAGTTACGAGGCGGAGCAGTGTGCGATGCCCTCGACGTGCGCGCGAA 497
Db 679 CACGCTGCGTGGCGGCCAGGACCTGGCGCGCTTGATCCACTGCTGTGCGACGCGAA 737

RESULT 43
CG268544/c 916 bp DNA linear GSS 25-AUG-2003
LOCUS CGX444TV.ZM.0.7.1.5.KB.Zea.mays.genomic.clone.ZM0626G16,
DEFINITION genomic survey sequence.
ACCESSION CG268544
VERSION CG268544.1 GI:34180685
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 916)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: CGX444TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..596
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/db_xref="taxon:39947"
/clone="BR030004000.PLATE_E01_5_005.ab1"
/tissue_type="roots"

FEATURES
source
1..596
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/db_xref="taxon:39947"
/clone="BR030004000.PLATE_E01_5_005.ab1"
/tissue_type="roots"

```

```

1..916
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM0626G16"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pCR4-TOPO; Site:1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 10.8%; Score 54.2; DB 29; Length 916;
Best Local Similarity 48.8%; Pred. No. 80;
Matches 146; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 199 TCCTTCGCCATGATCGTCGGCGTGTCCAGGCGAAGAGCAGCGAGTCCACCGG 258
Db 384 TCGGCTCTTACAGCGCAGCGCGGCCCTCGGCCCTTACCTTGAGCAGCAGC 325
QY 259 TCGGGCGTGAACCGCTCGCGGTTGCGACCGGATGTGCGCGGGTGTGAGCCGCCC 318
Db 324 TCCGCTTGGCACCCTTGTGCGTGTGAGCGGCTTGAGGTGCTTGACGGCGGCGCTTCCGGTCC 265
QY 319 TGCTTGGCGGCTGTGTCGACACCCAGGAGCAGGTCCGACCGATGCGGAGAG 378
Db 264 ACGGGGAACCTTGAAGACCCCGGACCATGAGTGCACCGCGCGTTCAGGGCCCTGGACG 205
QY 379 TTCGTACCGTGGCGCTTTCGCCGTGCGCGGTACGCCACACCGCTTGCCCTCGGCC 438
Db 204 TACACCATGACGGCGTTCGCCGTTCGCTCAGCGGAGCGCGGCCACCTCGTAC 145
QY 439 TTGAGCGAGTTACAGGCGGAGCAGTGTGCGATGCCCTCGACGTGCGCGCGAA 497
Db 144 CAGCGCTGCTGCGCGCGCAGGACCTGGCGCGCTTGATCCACTGCTGTGCGACGCGAA 86

RESULT 44
CG268544/c 596 bp mRNA linear EST 27-NOV-2002
LOCUS BR030004000.PLATE_E01_5_005.ab1.OA.Oryza.sativa (japonica
DEFINITION cultivar-group) cDNA clone BR030004000.PLATE_E01_5_005.ab1 similar
to contains ESTs AU097436 (S4951), D41930 (S4951), AU101944 (S2163),
AU095016 (E50186), AU058216 (E50186), C72995 (E2631),
D40291 (S2163), D41387 (S3866) unknown protein [Oryza sativa (japo,
mRNA sequence.
CA758758
CA758758.1 GI:25802797
EST.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 596)
Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M.,
Ferreira,H., Kawasaki,S., McColough,A., Michalowski,C.B.,
Palacio,C., Scara,G., Wheeler,M. and Zepeda,G.R.
Functional Genomics of Plant Stress Tolerance
Unpublished (2000)
Contact: Mark Fredrickson
Department of Plant Biology
University of Illinois
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 2172655473
Email: bohnertlab@life.uiuc.edu
Location/Qualifiers
1..596
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/db_xref="taxon:39947"
/clone="BR030004000.PLATE_E01_5_005.ab1"
/tissue_type="roots"

FEATURES
source
1..596
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/db_xref="taxon:39947"
/clone="BR030004000.PLATE_E01_5_005.ab1"
/tissue_type="roots"

```

```

/dev_stage="3-4 weeks"
/clone_lib="OA"
/note="19 h 200mM NaCl"

ORIGIN
Query Match      10.8%; Score 54; DB 14; Length 596;
Best Local Similarity 49.8%; Pred. No. 81;
Matches 135; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 189 GAATCTCTCTCTCTCGCATGATCTCTCGCGGTGTTCCAGGGAAGACGAGCGGTA 248
DB 366 GAACCCGCGTCTCTGAGCATGTTCCCTTGGCAGCGTTCGGGGGTACAGCGCCCTGAT 307

QY 249 GTTCACCGCGTTCGGCGGTGAACGCTCCGCGGTGCGACCGGGATGTGCGTTCGGCGGGT 308
DB 306 GCGCAGCGCGTTCGATGAGCGGCGCCGACCGCGGTCTCTCTCCACGCGCGGTGTGGAT 247

QY 309 GAGCGCGCGCTCTTGGCGCGGTCTGTGTCACACCGCAGGACCGAGTTCGGACCGAT 368
DB 246 GACGAGCGCCACCTGTCATCTTCCCTTGAACGCCCCACCGTACGAGTCCCAACCGTT 187

QY 369 GCGCGAGAAGTTGTCACGGTGGCGCTCTTCCCGGTTCGCGCGTACGCCACCGCCCGCTT 428
DB 186 GTCGCGTAGATGTTCTGATGGGAGCACCCTCCCTCGCGGGTCTACCGACACGTTGAG 127

QY 429 GCCTCGCGCTTACCGAGTTTCAGCAGGCG 459
DB 126 CTGCTCCGCTTGGCGCAGGTTCGCGCGCG 96

RESULT 45
LOCUS CG447961 866 bp DNA linear GSS 17-SEP-2003
DEFINITION OG9AA78TH.ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZMMBMA0849M11,
Genomic survey sequence.
ACCESSION CG447961
VERSION CG447961.1 GI:34832961
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 866)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
JOURNAL
COMMENT Other_GSSs: OG9AA78TV
Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
source
1..866
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0849M11"
/clone_lib="ZM 0.7 1.5 KB"
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      10.8%; Score 54; DB 29; Length 866;
Best Local Similarity 49.0%; Pred. No. 85;
Matches 144; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 189 GAATCTCTCTCTCTCGCATGATCTCTCGCGGTGTTCCAGGGAAGACGAGCGGTA 248
DB 366 GAACCCGCGTCTCTGAGCATGTTCCCTTGGCAGCGTTCGGGGGTACAGCGCCCTGAT 307

QY 249 GTTCACCGCGTTCGGCGGTGAACGCTCCGCGGTGCGACCGGGATGTGCGTTCGGCGGGT 308
DB 306 GCGCAGCGCGTTCGATGAGCGGCGCCGACCGCGGTCTCTCTCCACGCGCGGTGTGGAT 247

QY 309 GAGCGCGCGCTCTTGGCGCGGTCTGTGTCACACCGCAGGACCGAGTTCGGACCGAT 368
DB 246 GACGAGCGCCACCTGTCATCTTCCCTTGAACGCCCCACCGTACGAGTCCCAACCGTT 187

QY 369 GCGCGAGAAGTTGTCACGGTGGCGCTCTTCCCGGTTCGCGCGTACGCCACCGCCCGCTT 428
DB 186 GTCGCGTAGATGTTCTGATGGGAGCACCCTCCCTCGCGGGTCTACCGACACGTTGAG 127

QY 429 GCCTCGCGCTTACCGAGTTTCAGCAGGCG 459
DB 126 CTGCTCCGCTTGGCGCAGGTTCGCGCGCG 96

RESULT 46
LOCUS CG338271 851 bp DNA linear GSS 26-AUG-2003
DEFINITION OG0CG91TH.ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZMMBMA0685P13,
Genomic survey sequence.
ACCESSION CG338271
VERSION CG338271.1 GI:34255537
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 851)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
JOURNAL
COMMENT Other_GSSs: OG0CG91TV
Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
source
1..851
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0685P13"
/clone_lib="ZM 0.7 1.5 KB"
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      10.8%; Score 53.8; DB 29; Length 851;
Best Local Similarity 49.6%; Pred. No. 91;
Matches 170; Conservative 0; Mismatches 167; Indels 6; Gaps 1;

QY 109 GCGCGCGCGGACACCGCGGTGAGTTCGTCAGTGTACCGCGGGACGATACAGGATC 168
DB 23 GCGCTCCGCGCACCTCCGACCGCGGTGACCTCATCTGCACGCCCGCGCGAGGCTT 82

QY 169 CACTGTCCCGCAGCGCGCGGAATCTCTCTCTTCGCCATGATCTCTGCGCGGTTC 228
DB 83 CGTACCTTGAACCTCCCGCGGTGTAGCCCGCGGCCACACAGCTCGCGACTGCAGC 142

```

QY 229 CAGGCGAAGACGCGTAGTCCACCGCGTGGGCGTGAACGCGTCCGGGTGCGCACC 288  
 Db 143 ACAGGACACCATGCGCGCGCCACCGCTGTATACCGCGCGACGCGTC-----GGCACC 196  
 QY 289 GGGATGTGCGTCCGCGGGGAGACCGCCCTGCTTGGCCCGCGTGTGTGACACACCCAG 348  
 Db 197 TGGATGCGCGCCACACGAGGTGACGCGCCCTGTCGAGGCGACACAGACTCGGACCACTCGTCC 256  
 QY 349 GAGACGAGTCCGCGGACCGATGCGCAGAGATTCGTACGTTGGCGGTCTTCCCGTCCG 408  
 Db 257 GAGGACACACGAGTGCAGACCGCCGCTGTCGCCCGCGCGCGAGACCTCGAACGCGC 316  
 QY 409 CCGTACGCCACACCGCTTGGCCCTCGGCTTGAAGCGAGTTCA 451  
 Db 317 GCGTCTGTCGCGCCATCTCGGCGCTGACCGCGCAGCGCGTCCA 359

RESULT 47  
 AG171124/c 1798 bp DNA linear GSS 09-JAN-2002  
 LOCUS Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey  
 DEFINITION

ACCESSION AG171124 GI:16700802  
 VERSION AG171124  
 KEYWORDS GSS.  
 SOURCE Pan troglodytes (chimpanzee)  
 ORGANISM Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE 1  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 BAC end sequences of Library RPCI-43

UNPUBLISHED  
 2 (bases 1 to 1798)  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.

DIRECT SUBMISSION  
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:chimbases@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the chimpanzee BAC library RPCI-43. This BAC  
 end was generated during the R&D process and may have higher chance  
 of clone tracking errors.

PRIMERS

Sequencing: TJ

LIBRARY

Vector : pBACe3.6  
 R.Site 1 : EcoRI  
 R.Site 2 : EcoRI

Location/Qualifiers

1.1798  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="RP43-040F09.TJ"  
 /sex="male"  
 /cell\_type="lymphocytes"  
 /clone\_lib="RPCI-43 Chimpanzee Male BAC Library"

FEATURES

source

ORIGIN

Query Match 10.8%; Score 53.8; DB 29; Length 1798;  
 Best Local Similarity 42.9%; Pred. No. 99;  
 Matches 186; Conservative 0; Mismatches 247; Indels 1; Gaps 1;

QY 40 GCAGAGGAGAACCCCGGTATCGGGTAGACGTTGGCGCGGAGAGCGGTACGGCCCTT 99

Db 1134 GNGNGGCGGCCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075

QY 100 CGACGGTACG 159

QY 1074 CG 1015

QY 160 TACAGGATCCACTTCCGCGCAGCCCGCGCGAATCTCTCTCTCTCTCTCTCTCTCTCT 219

Db 1014 CG 955

QY 220 GCGTGGTTCAGGCGAAGAGCAGCGCGTAGTCCACCGCGTCCGGCGTGAACGCGTCCGCG 279

Db 954 GCGNNGCGCCGNG -CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 896

QY 280 GTGCGCACCGGAGATGTCGTGTCGCGCGGAGTGAAGCGCGCCCTGCTGTCGCGCGCGTGT 339

Db 895 GCG 836

QY 340 CACACCCAGGAGACAGGTCGCGGACCGATCCGCGCAGAAATTCGTACGCGTGGCGCTTCT 399  
 Db 935 CGCGNGCG 776  
 QY 400 GCCTCGCGCGCGTACGCCACACCGCTTCCCTTCGCGCTTGAGCGAGTTTCAGCAGGCG 459  
 Db 775 NCCG 716  
 QY 460 AGCAGTCCGTGCG 473  
 Db 715 CGCGCGCGCGCGCG 702

RESULT 48

CB659824/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CB659824 665 bp mRNA linear EST 09-APR-2003

OS:NE0401316.f OSJNED Oryza sativa (japonica cultivar-group) cDNA

clone OSJNED01B16 5', mRNA sequence.

CB659824

CB659824.1 GI:29663549

EST.

ORyza sativa (japonica cultivar-group)

Oriza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 665)

Jantaauriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,

Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.

Large-scale identification of ESTs involved in the interaction

between rice and Magnaporthe grisea

Unpublished (2003)

Contact: Rod Wing

Arizona Genomics Institute

University of Arizona

Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ

85721-0088, USA

Tel: 520 626 3967

Fax: 520 621 9288

Email: http://genome.arizona.edu

PCR Primers

FORWARD: gta aac cga cgg cca gtg

BACKWARD: gga aac agc tat gac cat g

Plate: 01 row: B column: 15

Seq primer: gta aac cga cgg cca gtg.

Location/Qualifiers

1.665

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="mRNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone="OSJNED01B16"

/tissue\_type="leaf"

/dev\_stage="3 week"

/lab\_host="DH10B"

/clone\_lib="OSJNED"

/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:

XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"



```
Db      370  GCCGTAACGCAATGCTCTCTGCTGGATTAAATCACACGCTGTAGATACTCTCGAAGCGCGACG 311
QY      245  CGTAGTCCACCGGTCGGGCGTGAAACGCGTCCGGGGTGGCCACCGGGATGTGGTGCCCG 304
Db      310  ACGACACGGGTGATGGAGGCGGTGGACCCGGAGGTGTGGGTGACGTGCACGGACATGAACC 251
QY      305  GGGTGAACCGGCGCTCTGCTTGCGCGGCGTCTGTCGCACACCCAGGAGACCCAGGTCCGGAC 364
Db      250  TGGTGGGAAGGCTTCCAGCTCGCGCTGCTGTGCACCAAGCCACCCCGGGACCGGC 191
QY      365  CGATGCCGCGAAGATTGCTCAGGTGGCGCTCTTGCGCGTGGCGGTAGCCACACCC 424
Db      190  CGACCATGCACGAGGTGCGAGGGTGTCTCTCCCTCTGCGCGCGCGCGCAAGCCTC 131
QY      425  GCTTGCCCTCGGCTTGAGCGAGTTCAGCAGGGCGAGCAGGTGCGGATGCCCTCGA 484
Db      130  CCGGCTCCAAAGGCGGCGCGGCTTCGGGCGGCGCGGCGACTACACGCGCTTCTCGCGA 71
QY      485  CGTGGCGG 493
Db      70  CGGCGGCG 62
```

Search completed: June 27, 2004, 20:50:43  
Job time : 2035.43 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 18:07:09 ; Search time 325.98 Seconds  
(without alignments)  
6516.033 Million cell updates/sec

Title: US-09-758-759-1\_COPY\_1\_500  
Perfect score: 500  
Sequence: 1 ggtaccgaccgtgtccgg.....tgcagtcggcggaacct 500

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues  
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001as:\*
- 5: Geneseq2001bs:\*
- 6: Geneseq2002s:\*
- 7: Geneseq2003as:\*
- 8: Geneseq2003bs:\*
- 9: Geneseq2003cs:\*
- 10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	500	100.0	37116	7	ABZ66810	Abz66810 Orthosomy
C 2	43972	83.8	13519	5	ABZ66785	Abz66785 Orthosomy
C 3	369	73.8	1251	7	ABZ66785	Abz66785 Orthosomy
C 4	298.4	59.7	10035	7	ABZ66813	Abz66813 Orthosomy
C 5	298.4	59.7	11115	4	ABL50562	ABL50562 Micromono
C 6	256.6	51.3	1251	7	ABZ66784	Abz66784 Orthosomy
C 7	149.8	30.0	1401	2	AAV34002	AAV34002 S. peucet
C 8	135	27.0	16020	3	AAA3283	Aaa3283 Streptomy
C 9	131	26.2	12152	5	AA508699	AA508699 M. carbon
C 10	124	24.8	603	7	ABZ66779	Abz66779 Orthosomy
C 11	118.2	23.6	15240	7	ABZ66809	Abz66809 Orthosomy
C 12	112.8	22.6	1281	7	ABZ66730	Abz66730 Orthosomy
C 13	97	19.4	42000	3	AAA63349	Aaa63349 Streptomy
C 14	95	19.4	63164	3	AAA63348	Aaa63348 Streptomy
C 15	95	19.0	25617	8	ACC58256	Acc58256 Novobioc
C 16	94	18.8	31248	8	ACC58250	Acc58250 Coumermyc
C 17	94	18.5	35359	8	ACC58251	Acc58251 Coumermyc
C 18	92.4	18.5	1290	7	ABZ66729	Abz66729 Orthosomy
C 19	92.4	18.5	14252	4	ABL50558	ABL50558 Micromono
C 20	92.4	18.5	48221	7	ABZ66811	Abz66811 Orthosomy
C 21	90.6	18.1	45055	7	ABZ66808	Abz66808 Orthosomy
C 22	87.8	17.6	1224	7	ABZ66728	Abz66728 Orthosomy
C 23	81	16.2	59616	7	ABZ37516	Abz37516 Streptomy

C 24	81	15.2	59816	7	ABZ37515	Abz37515 Streptomy
C 25	80.8	16.2	1224	7	ABZ37534	Abz37534 Streptomy
C 26	78	15.6	42291	8	ACC58253	ACC58253 Clorobloc
C 27	59.4	11.9	624	7	ABZ66778	Abz66778 Orthosomy
C 28	56.8	11.4	2122	2	AA70152	Aat70152 S. longisp
C 29	55.2	11.0	9994	4	AA85191	Aac85191 S. avermi
C 30	54.8	11.0	837	7	ACA27222	ACA27222 Prokaryot
C 31	54.6	10.9	9975	7	AA161173	AA161173 Actinosyn
C 32	54.6	10.9	82746	7	AA161224	AA161224 Actinosyn
C 33	54.4	10.9	3510	3	AA65259	AA65259 Maize sca
C 34	53.6	10.7	2964	7	ADA70431	Ada70431 Rice gene
C 35	53.4	10.7	1551	7	ACA37987	ACA37987 Prokaryot
C 36	53.4	10.7	27541	4	AA17185	AA17185 Streptomy
C 37	53.4	10.7	125401	4	AA17186	AA17186 Streptomy
C 38	53.2	10.6	2700	4	AA18767	AA18767 Human ANI
C 39	52.2	10.4	1200	7	ACA37556	ACA37556 Prokaryot
C 40	52	10.4	14055	7	AA161170	AA161170 Actinosyn
C 41	51.4	10.3	2742	7	AA161191	AA161191 Actinosyn
C 42	51.2	10.2	975	7	ABZ66676	ABZ66676 Orthosomy
C 43	51.2	10.2	2010	7	ACA23382	ACA23382 Prokaryot
C 44	50.6	10.1	546	2	AA76906	Aat76906 S. glauce
C 45	50.6	10.1	2697	7	ADA70563	Ada70563 Rice gene
C 46	50.6	10.1	88421	6	AA140781	AA140781 88421nt g
C 47	50.4	10.1	23666	2	AAQ10190	AAQ10190 Cephalosp
C 48	50.2	10.0	1248	7	ACA37895	ACA37895 Prokaryot
C 49	50.2	10.0	1605	2	AAx02914	Aax02914 Z. mays c
C 50	50.2	10.0	6733	2	AAx02913	Aax02913 Z. mays c

#### ALIGNMENTS

RESULT 1  
ABZ66810/c  
ID ABZ66810 standard; DNA; 37116 BP.

AC ABZ66810;

XX 27-OCT-2003 (revised)

DT 21-MAR-2003 (first entry)

XX

DE Orthosomycin biosynthetic gene cluster SEQ ID NO 279.

XX Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.

XX Micromonospora carbonacea; africana.

XX WO200279505-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-CA000432.

XX 28-MAR-2001; 2001US-0279095P.

XX 30-MAR-2001; 2001US-0279709P.

XX 20-APR-2001; 2001US-0285214P.

XX (ECOP-) ECOPIA BIOSCIENCES INC.

XX Farnet CM, Zazopoulos E, Staiffa A;

XX WPI; 2003-0584435/05.

XX Identifying orthosomycin biosynthetic gene, gene fragment or gene cluster, by detecting presence of nucleic acid sequence corresponding to 17 of flammomycins protein families.

XX Example 10; Page 458-478; 51lpp; English.

XX The invention relates to identifying orthosomycin biosynthetic genes and its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the presence of a nucleic acid sequence coding for a polypeptide (ABP99207-ABP99362). The method is useful for identifying an orthosomycin



CC biosynthetic gene, gene fragment or gene cluster, especially an  
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
CC gene fragment or gene cluster. The method is useful for detecting the  
CC presence of any organism that contains DNA for the production of  
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
CC orthosomycins) regardless of the level at which genes for orthosomycin  
CC production are expressed by the organism or the amount of orthosomycin  
CC produced by the organism. This allows for the detection of new  
CC orthosomycin natural products, not produced by the organism. (Updated on  
CC 27-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 37116 BP; 5171 A; 13140 C; 13619 G; 5186 T; 0 U; 0 Other;  
Query Match 100.0%; Score 500; DB 7; Length 37116;  
Best Local Similarity 100.0%; Pred. No. 9e-83;  
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGTACCCGACCGTCTCCGCGACACGAGTCCAGATACGCGGAGGACGACCCCGCGT 60  
Db 30486 GGTACCCGACCGTCTCCGCGACACGAGTCCAGATACGCGGAGGACACCCCGCGT 30427  
QY 61 AGTCGCGGTAGACGGTGGCGCGGAGGCGGTACGCGCCCTTCAGCGGTGAGCGGGCGGG 120  
Db 30426 AGTCGCGGTAGACGGTGGCGCGGAGGCGGTACGCGCCCTTCAGCGGTGAGCGGGCGGG 30367  
QY 121 ACACCGGGGGTCACTGTCACGTGTACGCGGGGAGCTACAGATCCACTGTCCGCCA 180  
Db 30366 ACACCGGGGGTCACTGTCACGTGTACGCGGGGAGCTACAGATCCACTGTCCGCCA 30307  
QY 181 GCCCGCGGAACTCCTGCTCTCTTCGCCATGATCTCGTCGGGTGTTCCAGCGGAAGGC 240  
Db 30306 GCCCGCGGAACTCCTGCTCTCTTCGCCATGATCTCGTCGGGTGTTCCAGCGGAAGGC 30247  
QY 241 AGCGGTGTAGTCACCGGTGCGGGCGTGAACGGGTCCGGGTGCGCACCGGATGTGGTG 300  
Db 30246 AGCGGTGTAGTCACCGGTGCGGGCGTGAACGGGTCCGGGTGCGCACCGGATGTGGTG 30187  
QY 301 CCGCGGGTACGCGCGCCCTGCTTGACCGCGGTGTCGCGCACCCAGGACACGAGTCC 360  
Db 30186 CCGCGGGTACGCGCGCCCTGCTTGACCGCGGTGTCGCGCACCCAGGACACGAGTCC 30127  
QY 361 GAACCGATGCCGAGAGTTCGTACGGTGGCGCTTCCTTCGCGCGCCGTACGCCACC 420  
Db 30126 GAACCGATGCCGAGAGTTCGTACGGTGGCGCTTCCTTCGCGCGCCGTACGCCACC 30067  
QY 421 ACCCGCTTGCCTCGCCCTTGAGCGAGTTCAGCAGCGGAGCAGGTGCGTGGATGCC 480  
Db 30066 ACCCGCTTGCCTCGCCCTTGAGCGAGTTCAGCAGCGGAGCAGGTGCGTGGATGCC 30007  
QY 481 TCGACGTGCGCGGAACT 500  
Db 30006 TCGACGTGCGCGGAACT 29987  
RESULT 2  
AAS08693  
ID AAS08693 standard; DNA; 109519 BP.  
XX AC  
XX AAS08693;  
XX  
XX 11-SEP-2003 (revised)  
DT 26-SEP-2003 (first entry)  
XX  
XX DE Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.  
XX Everninomycin; antibiotic; bottle-neck gene; orthomycin; fermentation;  
XX ds.  
XX  
XX OS Micromonospora sp. ATCC 39149.  
XX  
XX FH Location/Qualifiers  
FT complement(132..1382)  
FT CDS  
FT /\*tag= a

FT RBS  
FT complement(1389..1394)  
FT /\*tag= b  
FT CDS  
FT complement(1490..2611)  
FT /\*tag= c  
FT /\*product= "EvdB"  
FT complement(2618..2622)  
FT /\*tag= d  
FT complement(2622..3860)  
FT /\*tag= e  
FT /\*product= "EvdC"  
FT complement(3867..3870)  
FT /\*tag= f  
FT RBS  
FT 4134..4138  
FT /\*tag= h  
FT CDS  
FT 4143..5312  
FT /\*tag= g  
FT /\*product= "EvdD"  
FT 5309..6235  
FT /\*tag= i  
FT /\*product= "EvdE"  
FT 6226..6229  
FT /\*tag= k  
FT CDS  
FT 6232..7275  
FT /\*tag= j  
FT /\*product= "EvdF"  
FT 7272..8527  
FT /\*tag= l  
FT /\*product= "EvdG"  
FT 8333..8336  
FT /\*tag= n  
FT CDS  
FT 8342..9364  
FT /\*tag= m  
FT /\*product= "EvdH"  
FT complement(9463..10224)  
FT /\*tag= o  
FT /\*product= "EvdI"  
FT complement(10232..10235)  
FT /\*tag= p  
FT CDS  
FT 10424..11176  
FT /\*tag= q  
FT /\*product= "EvdJ"  
FT 12027..12455  
FT /\*tag= r  
FT /\*product= "EvdK"  
FT /\*note= "No start codon"  
FT complement(12108..13022)  
FT /\*tag= s  
FT /\*product= "EvdL"  
FT complement(13027..13030)  
FT /\*tag= t  
FT CDS  
FT complement(14410..15363)  
FT /\*tag= u  
FT /\*product= "EvrA"  
FT complement(15369..15373)  
FT /\*tag= v  
FT CDS  
FT complement(15380..16414)  
FT /\*tag= w  
FT /\*product= "EvrB"  
FT complement(16419..17873)  
FT /\*tag= x  
FT /\*product= "EvrC"  
FT complement(17870..18934)  
FT /\*tag= y  
FT /\*product= "EvrD"  
FT 19374..20906  
FT /\*tag= z  
FT /\*product= "EvrE"  
FT 21056..22542  
FT /\*tag= ab  
FT CDS  
FT 21064..22542  
FT /\*tag= aa

```
FT RBS /product= "EvrF" 22736. .22740
FT FT /*tag= ad 22748. .24172
FT CDS /*tag= ac 22748. .24172
FT CDS /product= "EvrG" complement(24177. .25223)
FT FT /*tag= ae 22748. .24172
FT RBS /product= "EvrH" complement(25230. .25233)
FT FT /*tag= af 22748. .24172
FT CDS 25550. .26626
FT FT /*tag= ag 22748. .24172
FT RBS /product= "EvrI" 26672. .26676
FT FT /*tag= ai 26685. .30479
FT CDS /*tag= ah 26685. .30479
FT FT /product= "EvrJ" complement(30557. .31876)
FT CDS /*tag= aj 26685. .30479
FT FT /product= "EvrK" complement(31885. .31888)
FT RBS /*tag= ak 26685. .30479
FT CDS complement(31941. .32882)
FT FT /*tag= al 26685. .30479
FT CDS /product= "EvrL" complement(33167. .34405)
FT FT /*tag= am 26685. .30479
FT RBS /product= "EvrM" complement(34414. .34418)
FT FT /*tag= an 26685. .30479
FT CDS complement(34449. .35210)
FT FT /*tag= ao 26685. .30479
FT RBS /product= "EvrN" complement(35219. .35221)
FT FT /*tag= ap 26685. .30479
FT CDS complement(35294. .36238)
FT FT /*tag= aq 26685. .30479
FT CDS /product= "EvrO" complement(36235. .36963)
FT FT /*tag= ar 26685. .30479
FT CDS /product= "EvrP" complement(36998. .38026)
FT FT /*tag= as 26685. .30479
FT CDS /product= "EvrQ" complement(38072. .38566)
FT FT /*tag= at 26685. .30479
FT CDS /product= "EvrR" complement(38892. .40163)
FT FT /*tag= au 26685. .30479
FT CDS /product= "EvrS" complement(40216. .40890)
FT FT /*tag= av 26685. .30479
FT CDS /product= "EvrT" complement(40887. .41576)
FT FT /*tag= ax 26685. .30479
FT RBS /product= "EvrU" complement(40899. .40902)
FT FT /*tag= aw 26685. .30479
FT CDS complement(41679. .42707)
FT FT /*tag= ay 26685. .30479
FT RBS /product= "EvrV" complement(42714. .42717)
FT FT /*tag= az 26685. .30479
FT CDS complement(42810. .43799)
FT FT /*tag= ba 26685. .30479
FT CDS /product= "EvrW" complement(43799. .44866)
FT FT /*tag= bc 26685. .30479
FT RBS /product= "EvrX" complement(43807. .43811)
FT FT /*tag= bb 26685. .30479

complement(45014. .45760)
/*tag= bd
/product= "EvrY"
complement(45767. .45770)
/*tag= be
complement(45952. .45956)
/*tag= bg
complement(45962. .46714)
/*tag= bf
/product= "EvrZ"
complement(47156. .49234)
/*tag= bh
/product= "EvrA"
51627. .52715
/*tag= bi
/product= "EvrB"
51629. .51622
/*tag= bj
52889. .53557
/*tag= bk
/product= "EvrC"
53554. .54207
/*tag= bl
/product= "EvrA"
complement(54362. .55117)
/*tag= bm
/product= "EvrB"
complement(55125. .55128)
/*tag= bn
complement(55135. .56094)
/*tag= bo
/product= "EvrC"
complement(56100. .56103)
/*tag= bp
complement(56184. .56813)
/*tag= bq
/product= "EvrC2"

Query Match 99.8%; Score 499.2; DB 5; Length 109519;
Best Local Similarity 99.6%; Pred. No. 1.1e-82;
Matches 498; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTACCCGACCGTGTCCCGGACACGAGTTCGAGTACGGGAGGAGGACACCCCGGT 60
DB 1 GGTACCCGACCGTGTCCCGGACACGAGTTCGAGTACGGGAGGAGGACACCCCGGT 60
QY 61 AGTCCCGGTAGACCGTGGCGCGGAGGCGTACCGGCTTCGACGGTTCAGCGGGCGGGCGG 120
DB 61 AGTCCCGGTAGACCGTGGCGCGGAGGCGTACCGGCTTCGACGGTTCAGCGGGCGGGCGG 120
QY 121 ACACCGCGGGTACGCTCGTACGTTACGCGGGGACGTACAGGATCCACTGTCGCCA 180
DB 121 ACACCGCGGGTACGCTCGTACGTTACGCGGGGACGTACAGGATCCACTGTCGCCA 180
QY 181 GCCCGGCGGAACCTCCTGCTCCTTCGCCATGATCTCGTGGGTGTTCCAGCGAGAGC 240
DB 181 GCCCGGCGGAACCTCCTGCTCCTTCGCCATGATCTCGTGGGTGTTCCAGCGAGAGC 240
QY 241 AGCGGTAGTCCACCGCGTTCGGGCGTGAACCGGTCGCGGGTTCGACCGGATGTCGTG 300
DB 241 AGCGGTAGTCCACCGCGTTCGGGCGTGAACCGGTCGCGGGTTCGACCGGATGTCGTG 300
QY 301 CGCGGGGTGAGCGCGGCGCTGCTTGGCGGCGTTCGTCGACACCCAGAGACAGGTCC 360
DB 301 CGCGGGGTGAGCGCGGCGCTGCTTGGCGGCGTTCGTCGACACCCAGAGACAGGTCC 360
QY 361 GGACCGATGCCGACAGAGTTCGTACGTTGCGGCTTTTCGCCGTCGCCGCGTACGCCACC 420
DB 361 GGACCGATGCCGACAGAGTTCGTACGTTGCGGCTTTTCGCCGTCGCCGCGTACGCCACC 420
QY 421 ACCCGCTTGCCTCGGCGCTTTCGAGGAGTTCAGCGGAGGAGGAGGAGGAGGAGGAGG 480
DB 421 ACCCGCTTGCCTCGGCGCTTTCGAGGAGTTCAGCGGAGGAGGAGGAGGAGGAGGAGG 480
```

```
QY 481 TCAGCTGCGCGCGGAACT 500
DB 481 TCAGCTGCGCGCGGAACT 500

RESULT 3
ABZ66785/c
ID ABZ66785 standard; DNA; 1251 BP.
XX
AC ABZ66785;
XX
DT 27-OCT-2003 (revised)
DT 21-MAR-2003 (first entry)
XX
DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 232.
XX
KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.
XX
OS Micromonospora carbonacea; africana.
XX
PN WO200279505-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-CA000432.
XX
PR 28-MAR-2001; 2001US-0279095P.
PR 30-MAR-2001; 2001US-0279709P.
PR 20-APR-2001; 2001US-0285214P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI Farnet CM, Zazopoulos E, Staffa A;
XX
DR WPI; 2003-058435/05.
DR P-FSDB; ABP99322.
XX
XX
Identifying orthosomycin biosynthetic gene, gene fragment or gene
cluster, by detecting presence of nucleic acid sequence corresponding to
17 of flambamycins protein families.
XX
PS Claim 16; Page 382; 51lpp; English.
XX
CC The invention relates to identifying orthosomycin biosynthetic genes and
its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the
presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
ABP99362). The method is useful for identifying an orthosomycin
biosynthetic gene, gene fragment or gene cluster, especially an
everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,
gene fragment or gene cluster. The method is useful for detecting the
presence of any organism that contains DNA for the production of
orthosomycins (both everninomicin-type orthosomycins and avilamycin-type
orthosomycins) regardless of the level at which genes for orthosomycin
production are expressed by the organism or the amount of orthosomycin
produced by the organism. This allows for the detection of new
orthosomycin natural products, not produced by the organism. (Updated on
27-OCT-2003 to standardise OS field)
XX
SQ Sequence 1251 BP; 181 A; 438 C; 439 G; 193 T; 0 U; 0 Other;
Query Match 73.8%; Score 369; DB 7; Length 1251;
Best Local Similarity 100.0%; Pred No. 9.6e-59;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 132 TCAGCTGCTGCTGCGCGGACGTACAGATCCATGTCGCCACGCCGCGGAA 191
DB 1251 TCAGCTGCTGCTGCGCGGACGTACAGATCCATGTCGCCACGCCGCGGAA 1192
QY 192 CTCTGCTCTCTGCGCATGATCTCGTCGGCGTTCAGGGCGAAGACGACGCGTAGNC 251
DB 1191 CTCTGCTCTCTGCGCATGATCTCGTCGGCGTTCAGGGCGAAGACGACGCGTAGTC 1132
```

```
QY 252 CACCGCGTCGGCGTGAAACGCGTCGCGGGTGCACACCGGATGTGCGTCCCGGGGTGAG 311
DB 1131 CACCGCGTCGGCGTGAAACGCGTCGCGGGTGCACACCGGATGTGCGTCCCGGGGTGAG 1072
QY 312 CCGGCCCTGCTTGGCGCGGTCGTGTCGCACACCCACAGGAGACGAGTCCGACCGGATGCC 371
DB 1071 CCGGCCCTGCTTGGCGCGGTCGTGTCGCACACCCACAGGAGACGAGTCCGACCGGATGCC 1012
QY 372 GCAGAGATTGCTCACGCTGCGCTCTTTCGCGCTGCGCGGTACGCCACACCGCTTGCC 431
DB 1011 GCAGAGATTGCTCACGCTGCGCTCTTTCGCGCTGCGCGGTACGCCACACCGCTTGCC 952
QY 432 CTGGCCCTTGCAGCGAGTTTCAGCAGGCGGAGCGAGTGGTGGATGCCCTCGACGTCGGC 491
DB 951 CTGGCCCTTGCAGCGAGTTTCAGCAGGCGGAGCGAGTGGTGGATGCCCTCGACGTCGGC 892
QY 492 GCGGAACT 500
DB 891 GCGGAACT 883

RESULT 4
ABZ66813/c
ID ABZ66813 standard; DNA; 10035 BP.
XX
AC ABZ66813;
XX
DT 27-OCT-2003 (revised)
DT 21-MAR-2003 (first entry)
XX
DE Orthosomycin biosynthetic gene cluster SEQ ID NO 282.
XX
KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.
XX
OS Micromonospora carbonacea; aurantiaca.
XX
PN WO200279505-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-CA000432.
XX
PR 28-MAR-2001; 2001US-0279095P.
PR 30-MAR-2001; 2001US-0279709P.
PR 20-APR-2001; 2001US-0285214P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI Farnet CM, Zazopoulos E, Staffa A;
XX
DR WPI; 2003-058435/05.
XX
Identifying orthosomycin biosynthetic gene, gene fragment or gene
cluster, by detecting presence of nucleic acid sequence corresponding to
17 of flambamycins protein families.
XX
PS Example 1; Page 506-511; 51lpp; English.
XX
CC The invention relates to identifying orthosomycin biosynthetic genes and
its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the
presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
ABP99362). The method is useful for identifying an orthosomycin
biosynthetic gene, gene fragment or gene cluster, especially an
everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,
gene fragment or gene cluster. The method is useful for detecting the
presence of any organism that contains DNA for the production of
orthosomycins (both everninomicin-type orthosomycins and avilamycin-type
orthosomycins) regardless of the level at which genes for orthosomycin
production are expressed by the organism or the amount of orthosomycin
produced by the organism. This allows for the detection of new
orthosomycin natural products, not produced by the organism. (Updated on
27-OCT-2003 to standardise OS field)
XX
```

SQ Sequence 10035 BP; 1396 A; 3516 C; 3657 G; 1466 T; 0 U; 0 Other;  
Query Match 59.7%; Score 298.4; DB 7; Length 10035;  
Best Local Similarity 76.1%; Pred. No. 6.1e-46;  
Matches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;  
QY 1 GGTACCCGACCGTCTCCGGACAAACGAGTCGAGATACGGGAGAGGACACCCCGGT 60  
DB 3742 GATATCCGAGCGTCTCGGTGAACACCGAGTCAGGTACGGGACACGAGACCCCGCG 3683  
QY 61 AGTCGGGTGAGCGTGGCGCGAAGCGGTACGCCCTTCAGCGGTACGCGGGGGGG 120  
DB 3682 GGTTCGGAAAGACCCGCGGGGTGAAGAGCTACGCCCTCGACGGCGAGTTCCGTGCTA 3623  
QY 121 ACACC---GGCGGTGAGCTGTCACGTGTACGGGGGACGTACAGGATCCACTGTCC 176  
DB 3622 CTGTACGGCCCACTCAATCCCGCACGTGAACGTGCGGAACGTACAGGATCCAGGCC 3563  
QY 177 GCCAGCCCGGGGAACTCTGTCTCTTCGCCATGATCTCGCGGTGTTCCAGGCGAA 236  
DB 3562 GCGGCGCTGCCGGAACGCGTCTCTTCGCCATGATCTCGCGGTGTTCCAGGCGAA 3503  
QY 237 GAGCAGCGGTAGTCCACCGGTTCGGGCGTGAAACCGGTTCGGGGTGCCACCGGATGTG 296  
DB 3502 GAGCAGCGGTAGTCCCGGGTTCGGTCCGGAACTCGTCCGCCGATGAACCGGGATGTG 3443  
QY 297 CGTGCGGGGTGAGCGGCGCTCTGTGCGGCGTGTGTCGACACCCAGGAGACAG 356  
DB 3442 CGTGCGGGGTGAGCGGCGCTCTGTGCGGCGTGTGTCGACACCCAGGAGACAG 3383  
QY 357 GTCCGGAACGATGCGCGAGAACTTCGTACGCGTGGCGTCTTTGCGCGTGGCGCGTACGC 416  
DB 3382 GTCCGCGCGATGCGCGAGAACTTCGTACGCGTGGCGTCTTTGCGCGTGGCGCGTACGC 3323  
QY 417 CACCAACCGCTTGCCTCGGCTTCGAGGAGTTACGAGGCGAGCAGGTGCGTGGCGAT 476  
DB 3322 CACCAACCGCTTGCCTCGGCTTCGAGGAGTTACGAGGCGAGCAGGTGCGTGGCGAT 3263  
QY 477 GCCCTGAGCTCGGCGGGAAC 498  
DB 3262 CCGGCGACCTGGGAGCGAAC 3241

## RESULT 5

ABL50562/c  
ID ABL50562 standard; DNA; 11115 BP.  
XX AC ABL50562;  
XX DT 18-JUN-2002 (first entry)  
XX DE Micromonospora carbonacea everninomicin locus nucleotide contig 8.  
XX KW Micromonospora carbonacea; antibiotic; everninomicin; biosynthesis;  
XX XW gene cluster; genetic manipulation; contig; gene; ds.  
XX OS Micromonospora carbonacea.

PN WO20015180-A2.

XX 02-AUG-2001.

XX 29-JAN-2001; 2001WO-CA000128.  
XX 27-JAN-2000; 2000US-017711P.XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX (FARN/) FARNET C.  
XX Stafira A, Zazopoulos E, Mercure S, Nowacki P;

XX WPI: 2001-476185/51.

DR P-PSDB; ABB06922, ABB06923, ABB06924, ABB06925, ABB06926, ABB06927,  
DR

DR ABB06928, ABB06929, ABB06930.

XX Novel isolated gene cluster encoding polypeptides involved in  
PT everninomicin biosynthesis useful for construction of everninomicin  
PT overproducing strains, and to allow chemical modifications of  
PT everninomicin to enhance certain properties.

PS Claim 1; Fig 1; 181pp; English.

XX ABL50555 to ABL50562 represent contigs 1 to 8 from the Micromonospora  
CC carbonacea everninomicin biosynthetic locus gene cluster. The contigs  
CC encode the protein sequences designated ORF (open reading frame) 1 to 49,  
CC given in ABB06981 to ABB06930. The gene cluster is useful for the  
CC construction of the everninomicin antibiotic in overproducing strains,  
CC and to allow chemical modifications of everninomicin to enhance certain  
CC properties via genetic manipulation or combination of biosynthesis. The  
CC gene cluster can be used to produce genetic systems and genes encoding  
CC novel enzyme activities, and avoid the problems of low yield and quality  
CC of everninomicins produced by chemical synthesis

XX SQ Sequence 11115 BP; 1521 A; 3924 C; 4057 G; 1613 T; 0 U; 0 Other;

Query Match 59.7%; Score 298.4; DB 4; Length 11115;  
Best Local Similarity 76.1%; Pred. No. 6.1e-46;  
Matches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;  
QY 1 GGTACCCGACCGTCTCCGGACAAACGAGTCGAGATACGGGAGAGGACACCCCGGT 60  
DB 3749 GATATCCGAGCGTCTCGGTGAACACCGAGTTCAGGTACGGGACACGAGACCCCGCG 3690  
QY 61 AGTCGGGTGAGCGTGGCGCGAAGCGGTACGCCCTTCAGCGGTGAGCGGGCGGGCGG 120  
DB 3689 GGTTCGGAAAGACCCGCGGGTGAAGACGTACGCCCTTCAGCGGGAGTTCGTTGCTA 3630  
QY 121 ACACC---GGCGGTGAGCTGTCACGTGTACGGGGGAGCTACAGGATCCACTGTCC 176  
DB 3629 CTTGACGGCCCCACATCCCGACGTGAACGTGCGGAACGTACAGGATCCAGCCCC 3570  
QY 177 GCCAGCCCGCGGAACCTCTGTCTCTTCGCCATGATCTCGTGGCGTGGTTCAGGCGAA 236  
DB 3569 GCGGCGCTGCCGGAACGCGTCTCTTCGCCATGATCTCGTGGCGTGGTTCAGCGAA 3510  
QY 237 GAGCAGCGGTAGTCCACCGGTTCGGGCGTGAACGCGTTCGGGGTCCGACCGGATGTG 296  
DB 3509 GAGCAGCGGTAGTCCCGGGGTGCGTTCGGGAACCTCGTCCGCGATGAACCGGGATGTG 3450  
QY 297 CGTGCGGGGTGAGCGGCGCTCTGTGCGCGGTGCTGTGCGACACCCAGGAGACCCAG 356  
DB 3449 CTTGCGGGGTGAGCGGCGCTCTGTGCGGGGTGCTGTGCTACACCCGCGACCCAG 3390  
QY 357 GTCCGGAACGATGCGCGAGAACTTCGTACGCGTGGCGTCTTTGCGCGTGGCGCGTACGC 416  
DB 3389 GTCCGCGCGATGCGCGAGAACTTCGCCACGGTTCGCGCTCTTTGGCGGTGCGCCCGTAGGC 3330  
QY 417 CACCAACCGCTTGCCTCGGCTTCAGGAGTTCAGCAGGCGGAGAGGTGCGTGGCGAT 476  
DB 3329 CACCAACCGCTTGCCTCGGCTTCAGGAGTTCAGCAGGCGGAGAGGTGCGTGGCGAT 3270  
QY 477 GCCCTGAGCTCGGCGGGAAC 498  
DB 3269 CCGGCGACCTGGGAGCGAAC 3248

## RESULT 6

ABZ66784/c  
ID ABZ66784 standard; DNA; 1251 BP.  
XX AC ABZ66784;  
XX DT 27-OCT-2003 (revised)  
XX DT 21-MAR-2003 (first entry)

DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 230.



QY 224 GGTTCAGGCGAAGAGCAGCGCGTAGTCCACCGCGTCCGGCGTGAACGCGTCCGGGGTGC 283  
 Db 1330 GGTTCAGGCGAAGAGCAGCGCATAGTCCGGGGTAGGGGGGCCCGGAATCCTCCGGCGAGC 1271  
 QY 284 GCACCGGGATGTGGTGC CGGGGGTGAAGCGCGCCCTCTTTGGCGGGTGGTGTGCGACA 343  
 Db 1270 GGATGGGGATGTGGGACCGGGGAGCGACGGCGGATCTTGGCGGGTGGTGTGCTAGA 1211  
 QY 344 CCCAGGAGACAGTCCCGGACCGATGCCGAGAAAGTTTCGTACGGTGGCGCTTTTCGCCG 403  
 Db 1210 CGCAGCGCAGCAGATCGGGGGCCGATCCCGCATAGTTGAGCACTGTGGCGCTCTTGCGG 1151  
 QY 404 TCGCGCGTACGCCACCAACCGCTTGGCTCGGCTTGGAGCGATTTCAGCAGGCGGAGCA 463  
 Db 1150 TCGCGCGTACCGACGACGATAGAGCGGAGTTCGCGACGTCGCGAGAGCGGCCACCA 1091  
 QY 464 GGTTCGGTCCGATCCCTTCGACGTCGCGCGGCGAACCT 500  
 Db 1090 GGTTCAGCAGACCGCTTGACCTGGCGCGGCACTT 1054

## RESULT 8

AAA39283/c  
 ID AAA39283 standard; DNA; 16020 BP.

XX AC  
 XX AC  
 XX AC

DT 08-SEP-2000 (first entry)

XX Streptomyces nogalater nogalamycin biosynthesis gene cluster SEQ ID NO:1.

XX Streptomyces nogalater; nogalamycin biosynthesis; antibiotic;  
 KW anthracycline biosynthetic pathway; gene cluster; drug screening;  
 KW antibiotic; antitumour antibiotic; anthracycline; ds.

XX Streptomyces nogalater.

XX WO200024775-A1.

XX 04-MAY-2000.

XX 20-OCT-1999; 99WO-FI000870.

XX 23-OCT-1999; 99FI-00002295.

XX (GALI-) GALILAEUS OY.

XX Ylibonko K, Torkkell S, Palmu K, Hakala J;

XX WPI; 2000-350696/30.

XX P-ESDB; AAY91055, AAY91056, AAY91057, AAY91058, AAY91059, AAY91060,  
 DR AAY91061, AAY91062, AAY91063, AAY91064, AAY91065, AAY91066, AAY91067,  
 DR AAY91068, AAY91069, AAY91070, AAY91071.

XX Isolated and purified DNA fragment for obtaining novel hybrid antibiotics  
 PT comprises the gene cluster for the anthracycline biosynthetic pathway of  
 PT the bacterium Streptomyces nogalater.

XX Claim 2; Page 32-40; 59pp; English.

XX The present sequence represents the nogalamycin biosynthesis gene cluster  
 CC isolated from Streptomyces nogalater. Nogalamycin is an anthracycline  
 CC antibiotic, so the nogalamycin biosynthetic pathway is also known as the  
 CC anthracycline biosynthetic pathway. DNA fragments, plasmids and process  
 CC from the present invention are useful for obtaining novel hybrid  
 CC antibiotics, such as anthracyclines (antitumour antibiotics) or  
 CC anthracyclonones

XX Sequence 16020 BP; 2295 A; 6011 C; 5425 G; 2189 T; 0 U; 100 Other;

Query Match

Best Local Similarity 27.0%; Score 135; DB 3; Length 16020;

61.5%; Pred. No. 4e-16;

Matches 216; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 QY 150 GCGGGGACGTACAGATCCACTGTCCCGCAGCCCGCGGAACCTCTCTCTTCGCCAT 209  
 Db 6323 GCGGGGATGGGACAGATGAACCGGCCCGCTTGTCCCGGAAGGCCGTCTCTCTTCGAG 6264  
 QY 210 GATCTCGTCCGCGTGTTCAGGGGAGAGCAGCGGTAGTCCACCGGTGGGGGTGAA 269  
 Db 6263 GATCTCCGTGGGTAGTTCAGGGGAGAGCAACAGTAGTAGTTCGGGGATGTTCCTTGGC 6204  
 QY 270 CGCGTCCGGGTGCGCACCGGATGTGCTGCCGGGGTGAAGCCGCCCTTGTTCGCCG 329  
 Db 6203 GTGCTCGGAGCGTCAACCGGTATGTGGGTGCGGGCAGCACCTTGCCTTCACTC 6144  
 QY 330 CGTCTGTGCGCACACCCAGGAGACAGTCCGGACCGATGCCGAGAGTTGTCACCGT 389  
 Db 6143 GGTGTGTGCTGTGAGTATTCAGTCTCTTCCAGCCGAGCCCGACACCGTGAAGTGT 6084  
 QY 390 GCGCTCTTCGCCGTGCGCGGTAGCCACACCGCTTGGCCCTTCGCGCTTGAAGCGATT 449  
 Db 6083 GTTGCCCTTGGCGGAGCACCGTAGCCGACGATGCGTTCGCCCTGGGCCACGAGGAGCG 6024  
 QY 450 CAGCAGGCGCAGCAGTGGTGGGATGCCCTCGACGTCCGGGGCGAACCT 500  
 Db 6023 CAGCAGTTCGCGCAGTTCGGTGGCCACCGCGTTCGATCCGCTGCGCGAACCT 5973

## RESULT 9

AAS08699  
 ID AAS08699 standard; DNA; 12152 BP.

XX AC  
 XX AAS08699;

XX 11-SEP-2003 (revised)

DT 26-SEP-2001 (first entry)

XX M. carbonacea DNA encoding Evernimycin biosynthetic enzymes.

XX Evernimycin; antibiotic; bottle-neck gene; orthomycin; fermentation;  
 KW ds.

XX Micromonospora sp. ATCC 39149.

XX Key Location/Qualifiers  
 FH complement(189..1064)

FT /tag= a

FT /product= "ORF1"

FT complement(1069..1073)

FT /tag= b

FT 1184..2767

FT /tag= c

FT /product= "ORF2"

FT complement(2683..3753)

FT /tag= d

FT /product= "ORF3"

FT complement(3766..4276)

FT /tag= e

FT /product= "ORF4"

FT complement(4280..4284)

FT /tag= f

FT complement(4526..5368)

FT /tag= g

FT /product= "ORF5"

FT complement(5392..6147)

FT /tag= h

FT /product= "ORF6"

FT complement(6152..6156)

FT /tag= i

FT complement(6194..7282)

FT /tag= j

FT /product= "ORF7"

FT complement(7280..8133)

FT /tag= k

```
FT RBS /product= "ORF8"  
FT complement (8141. .8145)  
FT *tag= l  
FT CDS complement (8254. .9318)  
FT *tag= m  
FT RBS /product= "ORF9"  
FT complement (9324. .9328)  
FT *tag= n  
FT RBS 9568. .9571  
FT *tag= p  
FT CDS 9575. .10504  
FT *tag= o  
FT /product= "ORF10"  
FT 10584. .11585  
FT *tag= q  
FT /product= "ORF11"  
XX  
PN W0200151639-A2.  
XX  
XX 19-JUL-2001.  
XX  
XX 12-JAN-2001; 2001WO-US001187.  
XX  
XX 12-JAN-2000; 2000US-0175751P.  
XX  
XX (SCHE ) SCHERING CORP.  
XX  
XX Hosted TJ, Horan AC, Wang TX;  
XX  
XX WPI; 2001-442147/47.  
XX P-PSDB; AAU04901, AAU04902, AAU04903, AAU04904, AAU04905, AAU04906,  
XX AAU04907, AAU04908, AAU04909, AAU04910, AAU04911.  
XX  
XX New nucleic acid molecules encoding everninomicin pathway gene products,  
XX useful for improving yields of everninomicin, to produce new  
XX everninomicin and as probes to identify homologous sequences.  
XX  
XX Example 1; Fig 12; 109pp; English.  
XX  
XX The sequence encodes 11 proteins comprising enzymes of the everninomicin  
XX antibiotic biosynthetic pathway. A vector comprising a M. carbonacea  
XX everninomicin biosynthetic pathway resistance gene product is useful for  
XX selecting for a transfected or transformed host cell. An integrative  
XX version of the vector is useful for introducing a everninomicin pathway  
XX gene (a bottle-neck gene) into an actinomycete of the genus  
XX Micromonospora. The DNA encoding the biosynthetic proteins is useful for  
XX synthesising novel everninomicin-related compounds, arising from  
XX modifications of the DNA sequence designed to change glycosyl and  
XX modified orsellinic acid groups contained in everninomicin, for  
XX expressing functional or mutant everninomicin biosynthetic enzyme for  
XX evaluation, diagnosis and preferably biosynthesis of everninomicin or  
XX other secondary metabolic products, improving the yield of everninomicins  
XX and to produce novel everninomicins and also as a hybridisation probe to  
XX identify homologous sequences. The encoded polypeptides are useful for  
XX combinatorial biosynthesis to generate libraries of orthomycins, e.g.  
XX everninomicin analogues/homologues and drug discovery. The DNA encoding  
XX the integrase allows for increasing a given gene dosage. The integrative  
XX vector can be used to permanently integrate copies of a heterologous gene  
XX of choice into chromosomes of different hosts and to integrate genes  
XX which increase the yield of known products or to generate novel products  
XX such as hybrid antibiotics or other novel secondary metabolites. The  
XX vector can also be used to integrate antibiotic resistance genes in order  
XX to carry out bioconversions with compounds to which the strain is  
XX normally sensitive and is thus useful in fermentation processes involving  
XX e.g. Streptomyces antibioticus. (Updated on 11-SEP-2003 to standardise OS  
XX field)  
XX  
XX Sequence 12152 BP; 1675 A; 4502 C; 4475 G; 1500 T; 0 U; 0 Other;  
XX  
XX Query Match 26.2%; Score 131; DB 5; Length 12152;  
XX Best Local Similarity 100.0%; Pred. No. 2.2e-15;  
XX Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 GGTACCCGACCGTGTCCCGGAACAACAGTTCGAGATACGGCGAGAGGAACACCCCCCGGT 60  
DB 12022 GGTACCCGACCGTGTCCCGGAACAACAGTTCGAGATACGGCGAGAGGAACACCCCCCGGT 12081  
QY 61 AGTCCGGGTAGACGGTGGGGCGGAGGGGTAGCCGCTTCGACGGTTCAGCGGGCGGGCGG 120  
DB 12082 AGTCCGGGTAGACGGTGGGGCGGAGGGGTAGCCGCTTCGACGGTTCAGCGGGCGGGCGG 12141  
QY 121 ACACCGCGCGGG 131  
DB 12142 ACACCGCGCGGG 12152  
XX  
XX RESULT 10  
XX ABZ66779/c  
XX ID ABZ66779 standard; DNA; 603 BP.  
XX  
XX AC ABZ66779;  
XX  
XX 27-OCT-2003 (revised)  
XX 21-MAR-2003 (first entry)  
XX  
XX Orthosomycin biosynthetic polynucleotide SEQ ID NO 220.  
XX  
XX Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.  
XX  
XX Micromonospora carbonacea; africana.  
XX  
XX W0200279505-A2.  
XX  
XX 10-OCT-2002.  
XX  
XX 28-MAR-2002; 2002WO-CA000432.  
XX  
XX 28-MAR-2001; 2001US-0279095P.  
XX 30-MAR-2001; 2001US-0279095P.  
XX 20-APR-2001; 2001US-0285214P.  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Farnet CM, Zazopoulos E, Staffa A;  
XX  
XX WPI; 2003-058435/05.  
XX P-PSDB; ABP99316.  
XX  
XX Identifying orthosomycin biosynthetic gene, gene fragment or gene  
XX cluster, by detecting presence of nucleic acid sequence corresponding to  
XX 17 of flamboyans protein families.  
XX  
XX Claim 16; Page 369-370; 51pp; English.  
XX  
XX The invention relates to identifying orthosomycin biosynthetic genes and  
XX its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
XX presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
XX ABP99362). The method is useful for identifying an orthosomycin  
XX biosynthetic gene, gene fragment or gene cluster, especially an  
XX everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
XX gene fragment or gene cluster. The method is useful for detecting the  
XX presence of any organism that contains DNA for the production of  
XX orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
XX orthosomycins) regardless of the level at which genes for orthosomycin  
XX production are expressed by the organism or the amount of orthosomycin  
XX produced by the organism. This allows for the detection of new  
XX orthosomycin natural products, not produced by the organism. (Updated on  
XX 27-OCT-2003 to standardise OS field)  
XX  
XX Sequence 603 BP; 77 A; 224 C; 197 G; 105 T; 0 U; 0 Other;  
XX  
XX Query Match 24.8%; Score 124; DB 7; Length 603;  
XX Best Local Similarity 100.0%; Pred. No. 5.3e-14;  
XX Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 GGTACCCGACCGTGTCCCGGAACAACAGTTCGAGATACGGCGAGAGGAACACCCCCCGGT 60
```

124 GGTACCCGACCGTGTCCCGGACACAGTCTGATACGGGAGAGAAACACCCCGGT 65  
61 AGTCCGGGTAGACGGTGGCGCGAAGCGGTACGGCCCTTCACCGGTAGCGGGCGCGG 120  
64 AGTCCGGGTAGACGGTGGCGCGAAGCGGTACGGCCCTTCACCGGTAGCGGGCGCGG 5  
121 ACAC 124  
4 ACAC 1  
RESULT 11  
ABZ66809/c  
ID ABZ66809 standard; DNA; 15240 BP.  
XX  
AC ABZ66809;  
XX  
DT 27-OCT-2003 (revised)  
DT 21-MAR-2003 (first entry)  
XX  
DE Orthosomycin biosynthetic gene cluster SEQ ID NO 278.  
XX  
KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.  
XX  
OS Micromonospora carbonacea; africana.  
XX  
PN WO200279505-A2.  
XX  
PD 10-OCT-2002.  
XX  
PF 28-MAR-2002; 2002WO-CA000432.  
XX  
PR 28-MAR-2001; 2001US-0279095P.  
PR 30-MAR-2001; 2001US-0279709P.  
PR 20-APR-2001; 2001US-0285214P.  
XX  
PA (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
PI Farnet CM, Zazopoulos E, Staffa A;  
XX  
DR WPI; 2003-058435/05.  
XX  
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene  
PT cluster, by detecting presence of nucleic acid sequence corresponding to  
PT 17 of flambamycins protein families.  
XX  
PS Example 10; Page 450-458; 51pp; English.  
XX  
CC The invention relates to identifying orthosomycin biosynthetic genes and  
CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
CC ABP99362). The method is useful for identifying an orthosomycin  
CC biosynthetic gene, gene fragment or gene cluster, especially an  
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic  
CC gene fragment or gene cluster. The method is useful for detecting the  
CC presence of any organism that contains DNA for the production of  
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
CC orthosomycins) regardless of the level at which genes for orthosomycin  
CC production are expressed by the organism or the amount of orthosomycin  
CC produced by the organism. This allows for the detection of new  
CC orthosomycin natural products, not produced by the organism. (Updated on  
CC 27-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 15240 BP; 2131 A; 5448 C; 5454 G; 2207 T; 0 U; 0 Other;  
Query Match 23.6%; Score 118.2; DB 7; Length 15240;  
Best Local Similarity 53.3%; Pred. No. 4.6e-13;  
Matches 249; Conservative 0; Mismatches 218; Indels 0; Gaps 0;  
33 AGATACCGGAGAGGACACCCCGGTAGTCCGGGTAGACGGTGGCGCGGAGGCGGTAC 92  
12684 AGAGATGATCCAGGGGAGACGACATGTCTCCGGGACCGGGCTTCGGGTGATGATGA 12625

93 GCGCTTTCACGGTACGGGCGGGCGGACACCGCGGGTACGCTGTCACGTGTACGG 152  
12624 AAGTTGATCACCTCGGTGTGGGGGGGTGTTCATCCGGTCCGCGACCGTCTCGACCG 12565  
153 GGGAGGTACAGATCACTGTCCGCGACCGCGGGGAACTCTGCTCTTTCGCCATGAT 212  
12564 GGGCAGCGGAAAGATCAACCCGGTGCCTGTGCGCAGCATCTCCGCTCCCGGGCGACGAT 12505  
213 CTCGTGCGGTGTCTCCAGGCGAAGAGCAGCGCTAGTCCACCGGTCCGGGGGTGAACGC 272  
12504 CTCGTCCCGAAGTGCAGGGCAGGACAGGTAGTAGTCCGGGCGGGCCCGCGGACTC 12445  
273 GTCGGGGTCCGACCGGATGTGCTCCGGGGTACCGCGCCCTGCTTGGCCGGCGGT 332  
12444 CTGCTCGCTGATGATTTTCGATGTCCGTGCGGAGGTGCGCGCCACCTTGTCCGGATT 12385  
333 CGTGTGCGACACCCAGGAGACAGGTCCGACCGATGCCAGAGTTCGTTCACGTGTGC 392  
12384 GCGTCCGGGGGTACCGGATGAGCTCGGGTCTATGCCGAGAACTGCAGGAGGTGTT 12325  
393 GCTCTTTCGCGCTCGCCCTACGCCACCGCGCTTGCCTCGGCTTTCGAGCGAGTTGAG 452  
12324 GCCCTTCGTGCGCGCGCTAGACGTGCACCGTGCAGCGCTGCGCCCGCGAGCTCGCGCAG 12265  
453 CAGGGCGAGCAGGTGCGGTGCGGATGCCCTCGAGCTCGGGCGGAACC 499  
12264 CAGGGCGCTCACCTGTCACGGTGTGGGCGACCTGCTCGGCGAAGC 12218  
RESULT 12  
ABZ66730/c  
ID ABZ66730 standard; DNA; 1281 BP.  
XX  
AC ABZ66730;  
XX  
DT 27-OCT-2003 (revised)  
DT 21-MAR-2003 (first entry)  
XX  
DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 122.  
XX  
KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.  
XX  
OS Micromonospora carbonacea; africana.  
XX  
PN WO200279505-A2.  
XX  
PD 10-OCT-2002.  
XX  
PF 28-MAR-2002; 2002WO-CA000432.  
XX  
PR 28-MAR-2001; 2001US-0279095P.  
PR 30-MAR-2001; 2001US-0279709P.  
PR 20-APR-2001; 2001US-0285214P.  
XX  
PA (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
PI Farnet CM, Zazopoulos E, Staffa A;  
XX  
DR WPI; 2003-058435/05.  
DR P-PSDB; ABP99267.  
XX  
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene  
PT cluster, by detecting presence of nucleic acid sequence corresponding to  
PT 17 of flambamycins protein families.  
XX  
PS Example 10; Page 275-276; 51pp; English.  
XX  
CC The invention relates to identifying orthosomycin biosynthetic genes and  
CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
CC ABP99362). The method is useful for identifying an orthosomycin  
CC biosynthetic gene, gene fragment or gene cluster, especially an



```
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,
CC gene fragment or gene cluster. The method is useful for detecting the
CC presence of any organism that contains DNA for the production of
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type
CC orthosomycins) regardless of the level at which genes for orthosomycin
CC production are expressed by the organism or the amount of orthosomycin
CC produced by the organism. This allows for the detection of new
CC orthosomycin natural products, not produced by the organism. (Updated on
CC 27-OCT-2003 to standardise OS field)
XX
SQ Sequence 1281 BP; 182 A; 440 C; 469 G; 190 T; 0 U; 0 Other;

Query Match      22.5%; Score 112.8; DB 7; Length 1281;
Best Local Similarity 56.5%; Pred. No. 5.5e-12;
Matches 210; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 128 CCGGTCAGCTCGTACGTGTACGCGGGGACGTACAGGATCCACTGTCGCGCAGCCCGC 187
Db 1276 CCGGTCGGCCGACCGCTCGACCCGCGGCGAGCGGAAGATCAACCGGTGCGCGTGGCCA 1217

QY 188 GGAATCTCTGCTTCGCGCATGATCTCTGCGGTGTTCCAGCGGAAGACGCGGT 247
Db 1216 GCATCTCCGCTCCCGGGGACGATCTCTCCGGAAGTGCCAGGGCAGGACCGGTAGT 1157

QY 248 AGTCACACCGCTCGGCGGTGAACCGCTCCGGGGTGCACACCGGGATGTCGTCGCGGGG 307
Db 1156 AGTCCGGCGGGCGCGCGCGACTCTGCTCGCTGATGATTTTCGATGTCGTCGCGAGG 1097

QY 308 TGAGCGCGCCCTGCTTGGCGCGGCTGCTGTGTCACACCGAGACCGAGTCCGACCGA 367
Db 1096 TGGCGCGCGCCACCTTGTCCGGATTGCGCTCCGCGCGTACCGGATGAGCTCGCGTCTA 1037

QY 368 TGCGCGCAGAAATTCGTCAGGTGGCGCTCTTCCGCTGCGCGGTACGCCACACCCCGCT 427
Db 1036 TGCGCGCAGAAATTCGAGAGGTTGTCCTTCTGTCGACGCGCGTAGAGTGCACCGTGC 977

QY 428 TGCCCTCGCGCTTGAGCGATTTCAGCAGGGCGAGCGTCCGTCGGATGCCCTCGACGT 487
Db 976 GGCCTCGGCGCGCGTCCGCGCAGCGCGCTCACCTCGTACCGTCTCGCGCACCT 917

QY 488 CGGCGCGGAC 499
Db 916 GCTCGCGGAGC 905

RESULT 13
AA63349
ID AAA63349 standard; DNA; 42000 BP.
XX
AC AAA63349;
XX
XX 06-MAR-2001 (first entry)
XX
XX Streptomyces globisporus C-1027 gene cluster ORF -7-25.
XX
XX Enediynes C-1027 biosynthesis gene cluster; apoprotein; chromophore;
XX cancer; ds.
XX
XX Streptomyces globisporus.
XX
XX Key Location/Qualifiers
XX CDS complement(8..658)
XX FT /*tag= a "ORF -7 protein"
XX FT complement(930..1478)
XX CDS /*tag= b
XX FT /product= "ORF -6 protein"
XX FT complement(1649..2713)
XX CDS /*tag= c
XX FT /product= "ORF -5 protein"
XX FT complement(2850..3237)
XX CDS /*tag= d
XX FT /product= "ORF -4 protein"
```

```
FT complement(3442..4971)
FT /*tag= e
FT /product= "ORF -3 protein"
CDS 5982..7479
FT /*tag= f
FT /product= "glycerol phosphate transporter"
CDS complement(7573..9900)
FT /*tag= g
FT /product= "ABC transport/UvrA-like protein"
CDS complement(9982..11349)
FT /*tag= h
FT /product= "Na+/H+ transporter"
CDS complement(11351..12835)
FT /*tag= i
FT /product= "hydroxylase/halogenase"
CDS 13012..14079
FT /*tag= j
FT /product= "dNDP-glucose synthase"
CDS complement(14212..14643)
FT /*tag= k
FT /product= "CagA"
CDS complement(14690..15922)
FT /*tag= l
FT /product= "aminotransferase"
CDS complement(15919..16653)
FT /*tag= m
FT /product= "N-methyl transferase"
CDS complement(16653..17924)
FT /*tag= n
FT /product= "C-methyl transferase"
CDS complement(18031..19191)
FT /*tag= o
FT /product= "epoxide hydrazine"
CDS complement(19267..19929)
FT /*tag= p
FT /product= "anthranilate synthase II"
CDS complement(19926..21407)
FT /*tag= q
FT /product= "anthranilate synthase I"
CDS complement(21424..22878)
FT /*tag= r
FT /product= "coenzyme F390 synthetase"
CDS complement(22875..23546)
FT /*tag= s
FT /product= "iron-sulphur flavoprotein"
CDS complement(23566..24702)
FT /*tag= t
FT /product= "O-acyl transferase"
CDS complement(24986..25564)
FT /*tag= u
FT /product= "epimerase"
CDS 25815..27170
FT /*tag= v
FT /product= "monooxygenase"
CDS 27214..28593
FT /*tag= w
FT /product= "glycosyl transferase"
CDS 28590..29588
FT /*tag= x
FT /product= "dNDP glucose dehydratase"
CDS 29632..31197
FT /*tag= y
FT /product= "SgsB transmembrane efflux protein"
CDS 31280..32590
FT /*tag= z
FT /product= "coenzyme F390 synthetase"
CDS 32809..34392
FT /*tag= aa
FT /product= "hydroxylase"
CDS complement(34458..35294)
FT /*tag= ab
FT /product= "ORF 22 protein"
FT /transl_except= (pos:35223..35225,aa:Ile)
```







```
FT CDS 22780..24042
FT /*tag= u
FT /gene= "novU"
FT /product= "NovU"
FT /note= "encodes ABR42576, C-methyltransferase"
FT 24104..24937
FT /*tag= v
FT /gene= "novV"
FT /product= "NovV"
FT /note= "encodes ABR42577, dTDP-1-glucose synthase"
FT 24994..25617
FT /*tag= w
FT /gene= "novW"
FT /product= "NovW"
FT /note= "encodes ABR42578, dTDP-4-keto-6-deoxyglucose 3,5-epimerase"
XX
XX PN W02003014352-A2.
XX
XX PD 20-FEB-2003.
XX
XX PF 06-AUG-2002; 2002WO-EP008777.
XX
XX PR 08-AUG-2001; 2001US-0310808P.
XX
XX PA (UYTU-) UNIV TUEBINGEN.
XX
XX PI Heide L, Li S;
XX
XX WPI; 2003-268202/26.
DR P-PSDB; ABR42556, ABR42557, ABR42558, ABR42559, ABR42560, ABR42561,
DR ABR42562, ABR42563, ABR42564, ABR42565, ABR42566, ABR42567, ABR42568,
DR ABR42569, ABR42570, ABR42571, ABR42572, ABR42573, ABR42574, ABR42575,
DR ABR42576, ABR42577, ABR42578.
XX
XX Novel polypeptide encoded by open reading frame of aminocoumarin
XX biosynthetic gene cluster, useful for synthesizing aminocoumarins useful
XX for treating bacterial infections and malignancies.
XX
XX PS Claim 2; Fig 4; 124pp; English.
XX
XX CC The present sequence is the complete sequence of the Streptomyces
XX sphaeroides novobioicin biosynthetic gene cluster. The sequences and
XX arrangement of the open reading frames of this gene cluster show
XX similarity to the biosynthetic gene clusters of the aminocoumarins
XX coumermycin A1, chlorobioicin and sinocyclinone (see ACC58250-53). The
XX invention provides aminocoumarin biosynthetic nucleic acids and
XX polypeptides, a method for producing hybrid antibiotics using a
XX combination of various regions of biosynthetic gene clusters, and a
XX pharmaceutical composition containing an aminocoumarin compound for use
XX in the treatment of Gram-positive bacterial infections and of malignant
XX diseases. (Updated on 27-OCT-2003 to standardise OS field)
XX
XX SQ Sequence 25617 BP; 4135 A; 8974 C; 8665 G; 3843 T; 0 U; 0 Other;
Query Match 19.0%; Score 95; DB 8; Length 25617;
Best Local Similarity 54.0%; Pred. No. 7.6e-09;
Matches 194; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
QY 140 TCACGTGTACCGGGGACGTACAGGATCCACTGTCCGCGAGCCCGCGGACCTCTGCT 199
DB 24019 TCACCCCGCGTTCGGGTATCGGAACGATGAATCGCGCGCGTTCCTCCAGGAATTCCTTCT 23960
QY 200 CCTTCGCGCATGATCTCGTCGCGTGTTCAGCGGAAGACGCGGTAGTCCACCGGT 259
DB 23959 CGTTGTCGATGATCTCGGTGCGTAGTTCCAGGCCAACACAGGTAGTAGTCCGTACAT 23900
QY 260 CGGCGGTGAACCGGTTCGGGGTGCACACCGGGATGTGCGTGCAGGGGTGAGCGGCGCT 319
DB 23899 GGGCGGCGCGTGTTCGGGTGACAGATCGGAATCTGGACCCCGCGAAGTATTTCCCT 23840
QY 320 GCTTGGCGGCGTCTGCTCCACACCCAGAGACCGGTCCGACCGATGCCGACAAT 379
|||||
```

```
Db 23839 GCTTGAGAACGGTTGTGTCTCGTGCAGAACCGGATCTCTGTATGTCCGAGCCCCGAGCGG 23780
QY 380 TCGTACAGGTGGGGCTCTTCGCGGTGCGCGGTACGCCACACCCGCTTGCCCTCGGCTT 439
Db 23779 CCAGCAGTGCCTGCGCTTGTGTCGGTGCCTGATCCCGCACGGTCTTGCGCTCGCCCA 23720
QY 440 TGAGCGAGTTACAGAGGGGAGCAGGTTCGTGGGATGCCCTCGACGTCCGCGCGGAC 498
Db 23719 CCAGCGCGGTACCAAGTTCCCGGATCTGTGCGCGCACCCCTTCGTACGATCCGCGGAC 23661
RESULT 16
ACC58250/c
ID ACC58250 standard; DNA; 31248 BP.
XX
XX AC ACC58250;
XX
XX DT 26-AUG-2003 (first entry)
XX
XX DE Coumermycin A1 biosynthetic gene cluster.
XX
XX KW Coumermycin A1; aminocoumarin; antibiotic; antibacterial; cytostatic;
XX gene; ds.
XX
XX OS Streptomyces rishiriensis.
XX
XX FH Key Location/Qualifiers
XX CDS 1..960
XX /*tag= a
XX /gene= "couG"
XX /note= "cuma", encodes ABR42475"
XX 1297..1512
XX /*tag= b
XX /gene= "couY"
XX /note= "similar to Mycobacterium tuberculosis mbtH,
XX encodes ABR42476"
XX 1818..3617
XX /*tag= c
XX /gene= "couH"
XX /product= "Acyl-CoA synthetase"
XX /note= "encodes ABR42477"
XX 3621..4844
XX /*tag= d
XX /gene= "couJ"
XX /product= "Cytochrome P450"
XX /note= "cumd, encodes ABR42478"
XX 5004..5780
XX /*tag= e
XX /gene= "couJ"
XX /product= "CouJ"
XX /note= "similar to 3-ketoacyl-(ACP)-reductase, encodes
XX ABR42479"
XX 5777..6514
XX /*tag= f
XX /gene= "couK"
XX /product= "Putative reductase"
XX /note= "encodes ABR42480"
XX 6521..8110
XX /*tag= g
XX /gene= "couL"
XX /product= "Putative ligase"
XX /note= "encodes ABR42481"
XX 8124..8332
XX /*tag= h
XX /gene= "couM"
XX /product= "Glycosyltransferase"
XX /note= "encodes ABR42482"
XX 9413..9700
XX /*tag= i
XX /gene= "couN1"
XX /product= "Unknown"
XX /note= "Cuml, encodes ABR42483"
XX 9711..10778
XX CDS
```

FT FT /\*tag= j  
FT FT /gene= "CouN2"  
FT FT /product= "3-Ketoacyl-(ACP)-synthase"  
FT FT /note= "CumJ, similar to Streptomyces peucetius DpsC,  
FT FT encodes ABR42484"  
FT FT 10775..11896  
FT FT /\*tag= k  
FT FT /gene= "CouN3"  
FT FT /product= "Acyl-CoA-dehydrogenase"  
FT FT /note= "encodes ABR42485"  
FT FT 12015..13520  
FT FT /\*tag= l  
FT FT /gene= "CouN4"  
FT FT /product= "Acyl-CoA-synthetase"  
FT FT /note= "proB, encodes ABR42486"  
FT FT 13524..13793  
FT FT /\*tag= m  
FT FT /gene= "CouN5"  
FT FT /product= "CouN5"  
FT FT /note= "similar to Pseudomonas fluorescens Pf-5 PtlL,  
FT FT encodes ABR42487"  
FT FT 13851..15533  
FT FT /\*tag= n  
FT FT /gene= "CouN6"  
FT FT /product= "Unknown"  
FT FT /note= "cumK, encodes ABR42488"  
FT FT 15582..16427  
FT FT /\*tag= o  
FT FT /gene= "CouN7"  
FT FT /product= "Putative hydrolase"  
FT FT /note= "cumL, encodes ABR42489"  
FT FT 16777..17469  
FT FT /\*tag= p  
FT FT /gene= "CouO"  
FT FT /product= "Putative methyltransferase"  
FT FT /note= "cumM, encodes ABR42490"  
FT FT 17869..18699  
FT FT /\*tag= q  
FT FT /gene= "CouP"  
FT FT /product= "O-methyltransferase"  
FT FT /note= "cumN, encodes ABR42491"  
FT FT 18796..20220  
FT FT /\*tag= r  
FT FT /gene= "CouR1"  
FT FT /product= "Unknown"  
FT FT /note= "cumO, encodes ABR42492"  
FT FT 20220..21353  
FT FT /\*tag= s  
FT FT /gene= "CouR2"  
FT FT /product= "Decarboxylase"  
FT FT /note= "cumP, encodes ABR42493, similar to  
FT FT phosphonopyruvate decarboxylase"  
FT FT 21386..22294  
FT FT /\*tag= t  
FT FT /gene= "CouR3"  
FT FT /product= "CouR3"  
FT FT /note= "cumO, encodes ABR42494, similar to Salmonella  
FT FT enterica serovar typhimurium PdxU"  
FT FT 22349..23518  
FT FT /\*tag= u  
FT FT /gene= "CouR4"  
FT FT /product= "Dehydrogenase"  
FT FT /note= "cumS, encodes ABR42495, similar to lactate and  
FT FT L(+)-mandelate dehydrogenases"  
FT FT complement(23572..25047)  
FT FT /\*tag= v  
FT FT /gene= "CouR5"  
FT FT /product= "Putative resistance protein"  
FT FT /note= "cumS, encodes ABR42496, transporter"  
FT FT 25185..26057  
FT FT /\*tag= w  
FT FT /gene= "CouR6"  
FT FT /product= "Transcript regulatory protein"

FT CDS /note= "cumT, encodes ABR42497"  
FT 26311..27177  
FT /\*tag= x  
FT /gene= "CouS"  
FT /product= "dTDP-4-keto-6-deoxyhexose reductase"  
FT /note= "cumU, encodes ABR42498"  
FT 27211..28221  
FT /\*tag= y  
FT /gene= "CouT"  
FT /product= "dTDP-glucose 4,6-dehydratase"  
FT /note= "cumV, encodes ABR42499"  
FT 28447..29709  
FT /\*tag= z  
FT /gene= "CouU"  
FT /product= "C-methyltransferase"  
FT /note= "cumW, encodes ABR42500"  
FT 29768..30658  
FT /\*tag= aa  
FT /gene= "CouV"  
FT /product= "dTDP-glucose synthase; glucose-1-phosphate  
FT thymidyltransferase"  
FT /note= "cumX, encodes ABR42501"  
FT 30655..31248  
FT /\*tag= ab  
FT /gene= "CouW"  
FT /product= "dTDP-4-keto-6-deoxyglucose 3,5-epimerase"  
FT /note= "cumY, encodes ABR42502, no stop codon"  
XX WO2003014352-A2.  
XX  
XX 20-FEB-2003.  
XX  
XX 06-AUG-2002; 2002WO-BP008777.  
XX  
XX 08-AUG-2001; 2001US-0310808P.  
XX (UYTU-) UNIV TUEBINGEN.  
XX Heide L, Li S;  
XX  
XX WPI; 2003-268202/26.  
XX P-PSDB; ABR42475, ABR42476, ABR42477, ABR42478, ABR42479, ABR42480,  
DR ABR42481, ABR42482, ABR42483, ABR42484, ABR42485, ABR42486, ABR42487,  
DR ABR42488, ABR42489, ABR42490, ABR42491, ABR42492, ABR42493, ABR42494,  
DR ABR42495, ABR42496, ABR42497, ABR42498, ABR42499, ABR42500, ABR42501,  
DR ABR42502.  
XX  
XX Novel polypeptide encoded by open reading frame of aminocoumarin  
PT biosynthetic gene cluster, useful for synthesizing aminocoumarins useful  
PT for treating bacterial infections and malignancies.  
XX Claim 2; Fig 1A; 124pp; English.  
XX  
XX The present sequence is a partial sequence of the Streptomyces  
CC rishtiensis strain DSM 40489 coumerycin A1 biosynthetic gene cluster.  
CC The sequence was isolated from a cosmid library using probes based on  
CC novobiocin-producer Streptomyces sphaeroides gene sequences. It includes  
CC 28 open reading frames (ORFs) that show marked sequence similarity to the  
CC corresponding ORFs of the novobiocin cluster, and in identical order. The  
CC invention provides aminocoumarin (e.g. coumerycin A1) biosynthetic  
CC nucleic acids and polypeptides, a method for producing hybrid antibiotics  
CC using a combination of various regions of biosynthetic gene clusters, and  
CC a pharmaceutical composition containing an aminocoumarin compound for use  
CC in the treatment of Gram-positive bacterial infections and of malignant  
CC diseases  
XX  
XX Query Match 18.8%; Score 94; DB 8; Length 31248;  
Best Local Similarity 53.9%; Pred. No. 1.1e-08;  
Matches 193; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
Qy 140 TCAGTGTACGCGGGAGCGTACAGGATCCACTGTCCGCCAGCCGCGGACTCCTGCT 199

```
Db 29686 TCACCGCGGTCGGGTATGGGAACATGAACCGCCGCCCTTCCAGGAAACACTTCT 29627
Qy 200 CTTCCGCATGATCTCGTCGGCGTGTTCAGAGGAGAGAGAGCGCGTAGTCCACCGCGT 259
Db 29626 CTTGGTGATGATCTCCAGCGGTAGTTCCAGGCCCAACAGCAGGTAGTAGTCCGGTAGCT 29567
Qy 260 CGGGCGTGACCGCTCCGGGTGCGCACCGGATGTCGTCGCGGGGTGAGCCGCGCT 319
Db 29566 GTCCTCTGCTGTTCGGCGACACAGATCGGACTCTCGAGCCCGTAGGAGTTTCCT 29507
Qy 320 GTTCCGCGCGGTCTGTGTCGACACCCAGGAGACAGATCCGAGCCGATGCGCGAGAAGT 379
Db 29506 GCTTCAGGCGCTGTTGTCGTCGACAGACCGGATCTCCTGATGCCAGCCCGATCGG 29447
Qy 380 TCGTCACGTGCGGCTCTTCGCGTCGCGCGTACGCCACACCGCGTTCGCTCGGCGCT 439
Db 29446 TCAGCAGTGCCTGCTTGTGTCGTCGTCGCGTACCGCTTCCAGCCACTGCTTTCGCTGCCA 29387
Qy 440 TGAGCGAGTTCACGAGGCGAGCAGGTGCGTCCGATGCCCTCGACGTGCGCGCGAA 497
Db 29386 CCAAGCTGGCCACCAAGTTCGCGATCTGTGCGCGCACCGCTTCCGTACGCTCGGCGAA 29329

RESULT 17
ACC58251/c
ID ACC58251 standard; DNA; 35359 BP.
XX
AC ACC58251;
XX
DT 26-AUG-2003 (first entry)
XX
DE Coumermycin A1 biosynthetic gene cluster.
XX
KW Coumermycin A1; aminocoumarin; antibiotic; antibacterial; cytostatic;
KW gene; ds.
XX
OS Streptomyces rishiriensis.
XX
FH Key Location/Qualifiers
CDS 1..1028
FT /*tag= a
FT /note= "ORF3, encodes ABR42503"
FT complement(1045..1536)
CDS
FT /*tag= b
FT /product= "putative FK506-binding protein"
FT /note= "encodes ABR42504"
FT 1535..1975
FT /*tag= c
FT /note= "ORF3, encodes ABR42505"
FT 2306..2959
FT /*tag= d
FT /gene= "CouE"
FT /product= "CouE"
FT /note= "encodes ABR42506"
FT 4112..5071
FT /*tag= e
FT /gene= "CouG"
FT /note= "cuma, encodes ABR42475"
FT 5408..5623
FT /*tag= f
FT /gene= "CouY"
FT /note= "similar to Mycobacterium tuberculosis mbtH,
FT encodes ABR42476"
CDS 5929..7728
FT /*tag= g
FT /gene= "CouH"
FT /product= "Acyl-CoA synthetase"
FT /note= "encodes ABR42477"
FT 7732..8955
FT /*tag= h
FT /gene= "CouI"
FT /product= "Cytochrome P450"
FT
```

```
FT /note= "cumD, encodes ABR42478"
FT 9115..9891
FT /*tag= i
FT /gene= "CouJ"
FT /product= "CouJ"
FT /note= "similar to 3-ketoacyl-(ACP)-reductase, encodes
FT ABR42479"
CDS 9888..10625
FT /*tag= j
FT /gene= "CouK"
FT /product= "Putative reductase"
FT /note= "encodes ABR42480"
FT 10632..12221
FT /*tag= k
FT /gene= "CouL"
FT /product= "Putative ligase"
FT /note= "encodes ABR42481"
FT 12235..13443
FT /*tag= l
FT /gene= "CouM"
FT /product= "Glycosyltransferase"
FT /note= "encodes ABR42482"
FT 13524..13811
FT /*tag= m
FT /gene= "CouN1"
FT /product= "Unknown"
FT /note= "CumI, encodes ABR42483"
FT 13822..14889
FT /*tag= n
FT /gene= "CouN2"
FT /product= "3-Ketoacyl-(ACP)-synthase"
FT /note= "CumJ, similar to Streptomyces peucetius DpsC,
FT encodes ABR42484"
CDS 14886..16007
FT /*tag= o
FT /gene= "CouN3"
FT /product= "Acyl-CoA-dehydrogenase"
FT /note= "encodes ABR42485"
FT 16126..17631
FT /*tag= p
FT /gene= "CouN4"
FT /product= "Acyl-CoA-synthetase"
FT /note= "proB, encodes ABR42486"
FT 17635..17904
FT /*tag= q
FT /gene= "CouN5"
FT /product= "CouN5"
FT /note= "similar to Pseudomonas fluorescens Pf-5 PtlL,
FT encodes ABR42487"
CDS 17962..19644
FT /*tag= r
FT /gene= "CouN6"
FT /product= "Unknown"
FT /note= "cumK, encodes ABR42488"
FT 19693..20538
FT /*tag= s
FT /gene= "CouN7"
FT /product= "Putative hydrolase"
FT /note= "cumL, encodes ABR42489"
FT 20888..21580
FT /*tag= t
FT /gene= "CouO"
FT /product= "Putative methyltransferase"
FT /note= "cumM, encodes ABR42490"
FT 21980..22810
FT /*tag= u
FT /gene= "CouP"
FT /product= "O-methyltransferase"
FT /note= "cumN, encodes ABR42491"
FT 22907..24331
FT /*tag= v
FT /gene= "CouR1"
FT /product= "Unknown"
FT
```





XX

XXI

55

U U

U U

1

oy

Qy

Ov

၁၃

73

**Dib**

2

2

RE

ID XX

XX

DT

DE  
W

KW KW

SO  
XY

XX  
DNI

XX

PD 10-OCT-2002.  
XX 28-MAR-2002; 2002WO-CA000432.  
XX 28-MAR-2001; 2001US-0279095P.  
XX 30-MAR-2001; 2001US-0279709P.  
XX 20-APR-2001; 2001US-0285214P.  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX Farnet CM, Zazopoulos E, Staffa A;  
XX WPI; 2003-058435/05.  
XX Identifying orthosomycin biosynthetic gene, gene fragment or gene  
XX cluster, by detecting presence of nucleic acid sequence corresponding to  
XX 17 of flambamycins protein families.  
XX Example 1; Page 479-505; 511pp; English.  
XX The invention relates to identifying orthosomycin biosynthetic genes and  
XX its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
XX presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
XX ABP99362). The method is useful for identifying an orthosomycin  
XX biosynthetic gene, gene fragment or gene cluster, especially an  
XX everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
XX gene fragment or gene cluster. The method is useful for detecting the  
XX presence of any organism that contains DNA for the production of  
XX orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
XX orthosomycins) regardless of the level at which genes for orthosomycin  
XX production are expressed by the organism or the amount of orthosomycin  
XX produced by the organism. This allows for the detection of new  
XX orthosomycin natural products, not produced by the organism. (Updated on  
XX 27-OCT-2003 to standardise OS field)  
XX Sequence 48221 BP; 7199 A; 16729 C; 16897 G; 7395 T; 0 U; 1 Other;  
Query Match 18.5%; Score 92.4; DB 7; Length 48221;  
Best Local Similarity 53.6%; Pred. No. 2.2e-08;  
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
QY 142 AGCTGTACGCGGGGACGTACAGGATCCACTGTCCGCGAGCCGCGGAACTCTCTGTCC 201  
DB 8244 ACCCGCAGCTGGGACGCGGGNAGATCAGTTGTTCCGCGCGCCACCGTGGCGCTCG 8195  
QY 202 TTCGCATGATCTCTCGCGGTGTTCCAGGCGAAGACAGCGGTAGTCCACCGCGTCG 261  
DB 8184 CCGCGCACGATCTCTCGTGAAGTGCCACGCGACGACGATGTTGTCGCGCGCGG 8125  
QY 262 GCGGTGAACGCTCGCGGTGCGCACCGGATGTCGTCGCGGGGTGAGCGCGCCCTGC 321  
DB 8124 GCCCGGAGTCCGCTCGCTGATGATCTCGATGTCGTTACGAGGTCCGCGCGCGAC 8065  
QY 322 TTGCGCGCGGTGCTGTCACACCCAGGAGACGAGTCCGACCGATCCCGCAGAGTTC 381  
DB 8064 TTGTCGCGGTTCGCTCGCGCGGTACGGATGACGCTGCGGTGATCCCGCAGTACTG 8005  
QY 382 GTACGCGGTGCTCTTCCGCTCGCGGTACGCGACACCGCTTCCGCTCGCGCTTG 441  
DB 8004 AGGAGGGTGTTCCTTGTGAGGCGCGGTAGAGTGCACGCTTCCGCTTCGCGGC 7945  
QY 442 AGCGAGTTCAGCAGGCGGAGCAGGTGCGGTGCGGATGCGCTCGATCCGCGCGCAACC 499  
DB 7944 AGACCATGAGCATCTTGACGAGTTGTCGCGGTGCGCGCGCCGACGCTTTCGCGCA 7887  
RESULT 21  
ID ABZ66808/c  
XX ABZ66808 standard; DNA; 45055 BP.  
AC ABZ66808;  
XX  
DB 21-MAR-2003 (first entry)

XX Orthosomycin biosynthetic gene cluster SEQ ID NO 277.  
XX Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.  
XX Streptomyces mobaraensis.  
XX WO200279505-A2.  
XX 10-OCT-2002.  
XX 28-MAR-2002; 2002WO-CA000432.  
XX 28-MAR-2001; 2001US-0279095P.  
XX 30-MAR-2001; 2001US-0279709P.  
XX 20-APR-2001; 2001US-0285214P.  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX Farnet CM, Zazopoulos E, Staffa A;  
XX WPI; 2003-058435/05.  
XX Identifying orthosomycin biosynthetic gene, gene fragment or gene  
XX cluster, by detecting presence of nucleic acid sequence corresponding to  
XX 17 of flambamycins protein families.  
XX Example 2; Page 425-450; 511pp; English.  
XX The invention relates to identifying orthosomycin biosynthetic genes and  
XX its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
XX presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
XX ABP99362). The method is useful for identifying an orthosomycin  
XX biosynthetic gene, gene fragment or gene cluster, especially an  
XX everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
XX gene fragment or gene cluster. The method is useful for detecting the  
XX presence of any organism that contains DNA for the production of  
XX orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
XX orthosomycins) regardless of the level at which genes for orthosomycin  
XX production are expressed by the organism or the amount of orthosomycin  
XX produced by the organism. This allows for the detection of new  
XX orthosomycin natural products, not produced by the organism  
XX Sequence 45055 BP; 6228 A; 16576 C; 16043 G; 6208 T; 0 U; 0 Other;  
Query Match 18.1%; Score 90.6; DB 7; Length 45055;  
Best Local Similarity 52.5%; Pred. No. 4.6e-08;  
Matches 198; Conservative 0; Mismatches 179; Indels 0; Gaps 0;  
QY 122 CACCGCGGTGAGTCTCGTACGCTGTACGCGGGGACGTACAGGATCCACTGTCCGCCAG 181  
DB 9974 CCCCGCACCGGTCAACACCTTCCAGGGGGGACGCGGAGACACAGTTCCGCCCG 9915  
QY 182 CCCCGCGGAACTCTCTCTTTCGCGATGATCTGTCGCGGTGTTTCCAGGGGAGAGCA 241  
DB 9914 CCGCCAGGAAGTCGCTTCCCGGGCGACCATCGCTCCGCTAGATCCACGGGAGACCA 9855  
QY 242 GCGCGTAGTCCACCGGTGCGGGGTGAACGCTGCGGGGTGCGCACCGGAGTGCCTGC 301  
DB 9854 GGAACCGGTCCGCTCCAGCGCCCGGCGCTCTTCTCGAGAGGATCGGAGTCTCGTGC 9795  
QY 302 CCGGGGTGAGCCCGCCCTCTGTTGGCGCGGTGCTGTCGACACCCAGGAGACAGATCCG 361  
DB 9794 CCGCGGTGAGGACGCGAACTTGTCTCTTACCTCGCGGTGTCACCGCAGGCGTCT 9735  
QY 362 GACCGATGCGCAGAGTTCGTTCACGCTGCGCTCTTCCCGTCCGCGGTAGCCACCA 421  
DB 9734 CGTCCAGGCGCAGTACTGAGGATACGTTGCCCTTGTGGAGGCGCGCTACCCGAGG 9675  
QY 422 CCGCGTTCGCTCGGCTTTCAGCGAGTTTCAGAGGCGGAGCAGGTGCGGTGCGGATCCCT 481  
DB 9674 TCGCGACCCCTTGTTCGCGGAGGTGTTGAGGAGTCCACGAGCCGCTCGCGGATTCCT 9615

QY 482 CGACGTCGGCGGAAC 498  
 Db 9614 CGGTCCGCGGCGGAAC 9598

## RESULT 22

ABZ66728/c  
 ID ABZ66728 standard; DNA; 1224 BP.

XX  
 AC ABZ66728;  
 DT  
 DT 21-MAR-2003 (first entry)  
 XX

DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 118.

XX Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.

XX Streptomyces mobaraensis.

OS

PN WO200279505-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-CA000432.

XX 28-MAR-2001; 2001US-0279095P.

XX 30-MAR-2001; 2001US-0279709P.

XX 20-APR-2001; 2001US-0285214P.

XX (ECOP-) ECOPIA BIOSCIENCES INC.

XX Farnet CW, Zazopoulos E, Staffa A;

XX WPI; 2003-058435/05.

XX P-PSDB; ABP99265.

XX Identifying orthosomycin biosynthetic gene, gene fragment or gene

XX cluster, by detecting presence of nucleic acid sequence corresponding to

XX 17 of flambamycins protein families.

XX Example 2; Page 271; 511pp; English.

XX The invention relates to identifying orthosomycin biosynthetic genes and

XX its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the

XX presence of a nucleic acid sequence coding for a polypeptide (ABP99207-

XX ABP99362). The method is useful for identifying an orthosomycin

XX biosynthetic gene, gene fragment or gene cluster, especially an

XX everninomicin-type or avilamycin-type orthosomycin biosynthetic

XX gene fragment or gene cluster. The method is useful for detecting the

XX presence of any organism that contains DNA for the production of

XX orthosomycins (both everninomicin-type orthosomycins and avilamycin-type

XX orthosomycins) regardless of the level at which genes for orthosomycin

XX production are expressed by the organism or the amount of orthosomycin

XX produced by the organism. This allows for the detection of new

XX orthosomycin natural products, not produced by the organism

XX

XX Sequence 1224 BP; 175 A; 474 C; 393 G; 182 T; 0 U; 0 Other;

XX

XX Query Match 17.6%; Score 87.8; DB 7; Length 1224;

XX Best Local Similarity 52.6%; Pred. No. 2e-07;

XX Matches 191; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

XX

XX 136 CTGTCACGTACGGGGGAGTACAGATCCAGTCCGACCGCGGGGAACTCC 195

XX 1223 CACACACATTCAGGGCGGGGACGAGGAGACAGGTTCGCGCGCGCGCGGAGAGTGC 1164

XX 196 TGGTCTCTTCGCCATGATCTCGTCGGGTGGTTCCAGGCGAAGAGACGCGTAGTCCACC 255

XX 1163 CGTTCGCGGCGGACCATCGGTCCCGGTAGATCCACGGGAGGACAGGAACCGTCCGCG 1104

XX 256 GCGTCGCGGTGACCGGTCCGGGTGCGGACCGGATGCGTGGCGGGGTGACCGCG 315

XX 1103 TCCAGGCGCGCGGCGCTCTTCTCGGAGACGATCGGATGTCGTGCGCGGCGTGAAGCAG 1044

QY 316 CCCTGCTTGGCCCGCGCTCGTGTGCGCACACCCAGGAGACAGGTTCGGACCCGATGCCGCGAG 375  
 Db 1043 CCGAATCTGCTCTGCTTACCTCGGCGATGCGACGGCAGAGCGTCTGCTCCAGGCGCGAG 984  
 QY 376 AAGTTGCTACGCTGCGGCTCTTTCGCGCTGCGCGCTGCGGCGTACGCCACACCCGCTTGCCTCG 435  
 Db 983 TACTGGAGGATGACGTTGGCTTGGTGGAGGCGCGCTACCCGAGGTGCGCAGCCCTTG 924  
 QY 436 GCCTTGAGCGAGTTCAGACAGGCGGACAGGTGCGTCCGATGCCCTCGACGTGCGGCGGCG 495  
 Db 923 TCCGCGAGGTGGTGAGGAGTCCACGAGCCGCTCGGGGATTCCTCGGTCGCGCGGCGG 864  
 QY 496 AAC 498  
 Db 863 AAC 861

## RESULT 23

ABZ37516

ID ABZ37516 standard; DNA; 59816 BP.

XX AC ABZ37516;

XX 26-FEB-2003 (first entry)

XX Streptomyces viridochromogenes Avi gene cluster sense strand.

XX Avilamycin; antibacterial; virucide; protozoicide; fungicide; infection;

XX medicine; Staphylococcus aureus; biosynthetic gene cluster; gene; ds.

XX Streptomyces viridochromogenes.

XX WO200268436-A1.

XX 06-SEP-2002.

XX 24-AUG-2001; 2001WO-BP009815.

XX 25-FEB-2001; 2001DE-01009166.

XX (COMB-) COMBINATURE BIOPHARM AG.

XX Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;

XX WPI; 2003-018650/01.

XX P-PSDB; ABP76680, ABP76681, ABP76682.

XX New avilamycin derivatives, useful for treatment of infections, and

XX nucleic acid encoding avilamycin synthesis enzymes.

XX Example 1; Page 68-301; 319pp; German.

XX The invention relates to avilamycin derivatives (I) with antibacterial,

XX virucide, protozoicide and fungicide activity. (I) are useful for

XX treatment of infections (bacterial, viral, protozoal or fungal), in human

XX or veterinary medicine, particularly where caused by Staphylococcus

XX aureus. (I) are more hydrophilic than known avilamycins. The present

XX sequence is that of a nucleic acid encoding avilamycin synthesis enzymes

XX from the Streptomyces viridochromogenes Avilamycin A biosynthetic gene

XX cluster (ABZ37515-ABZ37516)

XX Query Match 16.2%; Score 81; DB 7; Length 59816;

XX Best Local Similarity 50.9%; Pred. No. 2.5e-06;

XX Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

XX 122 CACCGCGCGGTGACGTCTGTCACGTGTACGGGGGAGTACAGGATCCACTGTCCGCGAG 181

XX 44797 CTCGCCACCCCGATCACACACTTCCAGCGTGGGCGGAGAACACAGCGTGCCTCGG 44856

XX 182 CCCGCGGGAATCTCTGCTCTTCTGCCATGATCTCGTGGCGGTGTTCTCCAGCGGAGACGA 241

Db 44857 AGGCCAGGAAGTCGGTTCCTCCGGCGACCAATCGCGTCCCGGTAGATCCACGGGAGGACCA 44916  
 QY 242 CGCGGTAGTCCACCGCGTCGGCGGTGAACGGGTCCGGGTGCGCACCGGGATGTGCGTGC 301  
 Db 44917 GGAAGTGGTGGGTCAAGCGCCCGGGCTCTCTCTCGGAGACGATCGGATGTGCGTGC 44976  
 QY 302 CGGGGTGAGCCGGCCCTGCTTGGCGCGGTGCTGTCGACACCGAGAGACAGGTCGCG 361  
 Db 44977 CGGGCGTGTAGACGCGCAACTTGTCTCTGTTCACTCGCGGATGACAGGCGAGAGCGTCT 45036  
 QY 362 GACCGATGCCGAGAGTTCGTACGGTGGCGGTCTTTCGCGGTGCGCGCGGTACGCCACCA 421  
 Db 45037 CGTCCAGCCGCGAGTACTGCAAGGATGACGTGCGCTCTCTCGAGCGCGCGTACCGAAGG 45096  
 QY 422 CCCGCTTGGCCCTCGCGCTTGAAGTTCAGCAGGCGAGAGTTCGCGGTGCGGTGCGGTGCGCT 481  
 Db 45097 TGTGACGCGCTTGTTCGCGCGAGCGGTGAGGAAATTCAGGAGTTCGCGTACGCGACTCTCT 45156  
 QY 482 CGACGTGGCGCGGCAAC 498  
 Db 45157 CGGTCCGCGCGGCAAC 45173  
 RESULT 24  
 ABZ37515/c  
 ID ABZ37515 standard; DNA; 59816 BP.  
 XX AC ABZ37515;  
 XX DT 05-MAR-2003 (first entry)  
 XX DE Streptomyces viridochromogenes Avi gene cluster sense strand.  
 XX KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
 XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; gene; ds.  
 XX OS Streptomyces viridochromogenes.  
 XX PN WO200268436-A1.  
 XX PD 06-SEP-2002.  
 XX PF 24-AUG-2001; 2001WO-EP009815.  
 XX PR 25-FEB-2001; 2001DE-01009166.  
 XX PA (COMB-) COMBINATURE BIOPHARM AG.  
 XX PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;  
 XX DR WPI; 2003-018650/01.  
 XX DR P-PSDB; ABB98398, ABP76679.  
 XX PT New avilamycin derivatives, useful for treatment of infections, and  
 XX PT nucleic acid encoding avilamycin synthesis enzymes.  
 XX PS Example 1; Page 68-301; 319pp; German.  
 XX CC The invention relates to avilamycin derivatives (I) with antibacterial,  
 CC virucide, protozoacide and fungicide activity. (I) are useful for  
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
 CC or veterinary medicine, particularly where caused by Staphylococcus  
 CC aureus. (I) are more hydrophilic than known avilamycins. The present  
 CC sequence is that of a nucleic acid encoding avilamycin synthesis  
 CC from the Streptomyces viridochromogenes Avilamycin A biosynthetic gene  
 CC cluster (ABZ37515-ABZ37516)  
 XX SQ Sequence 59816 BP; 8915 A; 21088 C; 21053 G; 8760 T; 0 U; 0 Other;  
 Query Match 16.2%; Score 81; DB 7; Length 59816;  
 Best Local Similarity 50.9%; Pred. No. 2.5e-06;  
 Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGGTCAGTCTGTCAGTGTACGCGGGGACGTACAGGATCCACTCTCGCGCCAG 181  
 Db 15020 CTTCGCGCAACCCGATCACACAATTCAGCGTGGCGAGCGGAAAGACCAAGGCTGCTCCGG 14961  
 QY 182 CCGCGCGGAACCTCTGCTTCCTTCGCAATGATCTCGTCGCGGTGGTTCCAGGGCGAAGACCA 241  
 Db 14960 AGGCCAGGAAGTTCGGTTCCTCCGGGCGACCATCGCTCCGGGTAGATCCACGGGAGACCA 14901  
 QY 242 GCGCGTAGTCCACCGCGGTGGCGGTGAACGGGTCCGGGTGCGCACCGGGATGTGCGTGC 301  
 Db 14900 GGAACCTGCTCGGGCTCAAGCGCGCGGCTCTCTCTCGAGACGATCGGATGTTCGTGC 14841  
 QY 302 CGGGGTGAGCGCGCGCTCTTGGCGCGGTGCTGCGCACACCGAGAGACCAAGTCCG 361  
 Db 14840 CGGGCGTGTAGCAGCGCAACTTGTCTCTGTTCACTCGCGATGAGGGCAGGAGCTCT 14781  
 QY 362 GACCGATGCCGAGAGTTCGTACGGTGGCGCTCTTTCGCGGTGCGGTGCGGTGCGGTGCGGT 421  
 Db 14780 CGTCCAGCGCGCAGTACTGCAAGTACGTTGCTTCTGCGAGCGCGGTAGCAGGAGT 14721  
 QY 422 CGCGTTCGCTCGCGCTTGGAGGATTCAGCAGGCGGAGCGGTGCGGTGCGGTGCGGTGCGGT 481  
 Db 14720 TGTGACGCGCTTGTTCGCGCGAGCGGTGAGGAAATTCAGGAGTTCGCGGTGCGGTGCGGT 14661  
 QY 482 CGACGTGGCGCGGCAAC 498  
 Db 14660 CGGTCCGCGCGGCAAC 14644  
 RESULT 25  
 ABZ37534/c  
 ID ABZ37534 standard; DNA; 1224 BP.  
 XX AC ABZ37534;  
 XX DT 26-FEB-2003 (first entry)  
 XX DE Streptomyces viridochromogenes AviI encoding polynucleotide.  
 XX KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
 XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; gene; ds.  
 XX OS Streptomyces viridochromogenes.  
 XX PN WO200268436-A1.  
 XX PD 06-SEP-2002.  
 XX PF 24-AUG-2001; 2001WO-EP009815.  
 XX PR 25-FEB-2001; 2001DE-01009166.  
 XX PA (COMB-) COMBINATURE BIOPHARM AG.  
 XX PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;  
 XX DR WPI; 2003-018650/01.  
 XX DR P-PSDB; ABP76700.  
 XX PT New avilamycin derivatives, useful for treatment of infections, and  
 XX PT nucleic acid encoding avilamycin synthesis enzymes.  
 XX PS Claim 13; Page 121-126; 319pp; German.  
 XX CC The invention relates to avilamycin derivatives (I) with antibacterial,  
 CC virucide, protozoacide and fungicide activity. (I) are useful for  
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
 CC or veterinary medicine, particularly where caused by Staphylococcus  
 CC aureus. (I) are more hydrophilic than known avilamycins. The present  
 CC sequence is that of a nucleic acid encoding avilamycin synthesis  
 CC from the Streptomyces viridochromogenes Avilamycin A biosynthetic gene  
 CC cluster (ABZ37515-ABZ37516)

XX SQ Sequence 1224 BP; 179 A; 456 C; 396 G; 193 T; 0 U; 0 Other;  
Query Match 16.2%; Score 80.8; DB 7; Length 1224;  
Best Local Similarity 53.1%; Pred. No. 3.8e-06;  
Matches 172; Conservative 0; Mismatches 152; Indels 0; Gaps 0;  
QY 175 CCACAGCCGGCGGAACCTCTCTCTCGCCATGATCTCGCGCGTGTTCAGGGCG 234  
DB 1184 CCTCCGAGGCCAGGAAGTCGGTTCCTCCGGCGACCATCGCGTAGATCCACGGG 1125  
QY 235 AAGAGAGCGCGTAGTCCACCGCGTCGGGGGTGAACCGTCCGGGGTGCACCGGGATG 294  
DB 1124 AGACAGGAACCTGTCGGCTCAAGCGCCCGGCTCTCTCGGAGACGATCGGGATG 1065  
QY 295 TCGTCCCGGGGTGAGCGCGCTCTGTGCGCGGTGTGTCGACACCCAGGAGACC 354  
DB 1064 TTCTGCGCGGCTGTAGACCGAAGTCTCTCTCTTACCTCGCGATGAGGGCAGG 1005  
QY 355 AGTCCGAGCCGATGCGCAGAAAGTTCGTACCGTGGCGCTTTCCTCGCGCGGTAC 414  
DB 1004 AGCGTCTGTCGACGCGCGAGTACTGACGAGTACGCTTGCCCTTCGTCGAGCGCGGTAG 945  
QY 415 GCCACACCGCTTGGCTCGGCTTGAGCGAGTTCAGCAGGCGGAGCGGTGCGG 474  
DB 944 CCAGGGTGTGACGCCCTTGTCCCGGAGGCGGTGAGGAATTCAGGAGTGTGTCACGG 885  
QY 475 ATGCCCTCGACGTCGGCGCGAAC 498  
DB 884 GACTCTCGTCCGCGCGCGAAC 861  
RESULT 26  
ID ACC58253/c  
XX ACC58253 standard; DNA; 42291 BP.  
XX AC ACC58253;  
XX DT 27-OCT-2003 (revised)  
XX DT 26-AUG-2003 (first entry)  
XX DE Clorobiocin biosynthetic gene cluster.  
XX KW Clorobiocin; aminocoumarin; antibiotic; antibacterial; cytostatic; gene;  
XX KW ds.  
XX OS Streptomyces roseochromogenes subsp. oscitans.  
XX FH Key Location/Qualifiers  
XX CDS 18..938  
FT /tag= a  
FT /note= "ORF1, encodes ABR42509, similar to Streptomyces  
FT hygroscopicus FkBL"  
FT complement(1020..1613)  
FT /tag= b  
FT /product= "sarcosine oxidase gamma subunit"  
FT /note= "ORF2, encodes ABR42510"  
FT complement(1642..4530)  
FT /tag= c  
FT /product= "sarcosine oxidase alpha subunit"  
FT /note= "ORF3, encodes ABR42511"  
FT complement(4527..4808)  
FT /tag= d  
FT /product= "sarcosine oxidase delta subunit"  
FT /note= "ORF4, encodes ABR42512"  
FT complement(4819..6039)  
FT /tag= e  
FT /product= "sarcosine oxidase beta subunit"  
FT /note= "ORF5, encodes ABR42513"  
FT complement(6036..7256)  
FT /tag= f  
FT /product= "Unknown"  
FT /note= "ORF6, encodes ABR42514, similar to serine

FT hydroxymethyltransferase"  
FT 7463..8119  
FT /tag= g  
FT /product= "Unknown"  
FT /note= "ORF7, encodes ABR42515, similar to gnt-R family  
FT transcriptional regulator"  
FT 8265..8714  
FT /tag= h  
FT /product= "Hypothetical protein"  
FT /note= "ORF8, encodes ABR42516"  
FT 8963..9199  
FT /tag= i  
FT /product= "Putative transposase"  
FT /note= "ORF9, encodes ABR42517"  
FT 9348..10001  
FT /tag= j  
FT /gene= "CioE"  
FT /product= "CioE"  
FT /note= "encodes ABR42518, similar to novE of novobiocin  
FT biosynthetic gene cluster"  
FT 10078..11166  
FT /tag= k  
FT /gene= "CioF"  
FT /product= "Putative oxidoreductase"  
FT /note= "encodes ABR42519, similar to novP of novobiocin  
FT biosynthetic gene cluster"  
FT 11746..12705  
FT /tag= l  
FT /gene= "CioG"  
FT /product= "Regulatory protein clog"  
FT /note= "encodes ABR42520"  
FT 12920..13135  
FT /tag= m  
FT /gene= "CioY"  
FT /product= "mbtH-like protein"  
FT /note= "encodes ABR42521"  
FT 13390..15192  
FT /tag= n  
FT /gene= "CioH"  
FT /product= "Peptide synthetase-like protein"  
FT /note= "encodes ABR42522"  
FT 15196..16419  
FT /tag= o  
FT /gene= "CioI"  
FT /product= "Cytochrome P450"  
FT /note= "encodes ABR42523"  
FT 16573..17349  
FT /tag= p  
FT /gene= "CioJ"  
FT /product= "Putative 3-oxoacyl-(ACP)-reductase"  
FT /note= "encodes ABR42524"  
FT 17346..18083  
FT /tag= q  
FT /gene= "CioK"  
FT /product= "Putative oxidoreductase"  
FT /note= "encodes ABR42525"  
FT 18090..19673  
FT /tag= x  
FT /gene= "CioL"  
FT /product= "Amide synthetase"  
FT /note= "similar to novobiocin acid synthetase, encodes  
FT ABR42526"  
FT 19687..20859  
FT /tag= s  
FT /gene= "CioM"  
FT /product= "Glycosyltransferase"  
FT /note= "encodes ABR42527"  
FT 20853..21140  
FT /tag= t  
FT /gene= "CioN1"  
FT /product= "Hypothetical protein"  
FT /note= "encodes ABR42528"  
FT 21150..22217  
FT CDS

FT FT /tag= u  
FT FT /gene= "clon2"  
FT FT /product= "Putative 3-ketoacyl- (ACP) -synthetase"  
FT FT /note= "encodes ABR42529"  
FT FT 22207. .23337  
FT FT /tag= u  
FT FT /gene= "clon3"  
FT FT /product= "Putative acyl-CoA dehydrogenase"  
FT FT /note= "encodes ABR42530"  
FT FT 23456. .24961  
FT FT /tag= v  
FT FT /gene= "clon4"  
FT FT /product= "Acyl-CoA synthetase"  
FT FT /note= "encodes ABR42531"  
FT FT 24965. .25234  
FT FT /tag= w  
FT FT /gene= "clon5"  
FT FT /product= "Clon5methyltransferase"  
FT FT /note= "encodes ABR42532, similar to acyl carrier  
FT FT protein"  
FT FT 25232. .26977  
FT FT /tag= x  
FT FT /gene= "clon6"  
FT FT /product= "Clon6", similar to coun6 of coumermycin  
FT FT A1"  
FT FT /note= "encodes ABR42533, similar to coun6 of coumermycin  
FT FT 27026. .27862  
FT FT /tag= y  
FT FT /gene= "clon7"  
FT FT /product= "Putative hydrolase"  
FT FT /note= "encodes ABR42534"  
FT FT 28099. .29673  
FT FT /tag= z  
FT FT /gene= "clon8"  
FT FT /product= "Putative halogenase"  
FT FT /note= "encodes ABR42535"  
FT FT 29998. .30831  
FT FT /tag= aa  
FT FT /gene= "clon9"  
FT FT /product= "O-methyltransferase"  
FT FT /note= "encodes ABR42536"  
FT FT 31832. .32806  
FT FT /tag= ab  
FT FT /gene= "clon10"  
FT FT /product= "Clon10"  
FT FT /note= "encodes ABR42537, similar to novQ of novobiocin  
FT FT biosynthetic gene cluster"  
FT FT 32803. .33636  
FT FT /tag= ac  
FT FT /gene= "clon11"  
FT FT /product= "Putative aldolase"  
FT FT /note= "encodes ABR42538"  
FT FT 34466. .35332  
FT FT /tag= ad  
FT FT /gene= "clon12"  
FT FT /product= "dtp-4-keto-6-deoxyhexose reductase"  
FT FT /note= "encodes ABR42539"  
FT FT 35366. .36376  
FT FT /tag= ae  
FT FT /gene= "clon13"  
FT FT /product= "dtp-glucose 4,6-dehydratase"  
FT FT /note= "encodes ABR42540"  
FT FT 36602. .37864  
FT FT /tag= af  
FT FT /gene= "clon14"  
FT FT /product= "C-methyltransferase"  
FT FT /note= "encodes ABR42541"  
FT FT 37976. .38866  
FT FT /tag= ag  
FT FT /gene= "clon15"  
FT FT /product= "dtp-glucose synthase"  
FT FT /note= "encodes ABR42542"  
FT FT 38863. .39459  
FT FT /tag= ah  
FT FT /gene= "clon16"

FT FT /product= "dtp-4-keto-6-deoxyglucose 3,5-epimerase"  
FT FT /note= "encodes ABR42543"  
FT FT 39459. .40220  
FT FT /tag= ai  
FT FT /gene= "clon17"  
FT FT /product= "Unknown"  
FT FT /note= "encodes ABR42544"  
FT FT 40574. .42291  
FT FT /tag= aj  
FT FT /gene= "gyrB"  
FT FT /partial  
FT FT /product= "Chlorobiocin-resistance gyrase B"  
FT FT /note= "encodes ABR42545, no stop codon"  
XX XX  
XX XX WO2003014352-A2.

Query Match 15.6%; Score 78; DB 8; Length 42291;  
Best Local Similarity 51.1%; Pred. No. 9.2e-06;  
Matches 183; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 140 TCACGTGTACGGGGGACGTACAGGATCCACTGTCCGCCACCGCGGCGAACTCTGCT 199  
DB 37841 TCACGGGTGTTTCGGGTATGGGAACAATAAAACGGCGCGCTTTTCCAGAAACGACTTCT 37782  
QY 200 CCTTCGCCATGATCTCGTCGGGTGTTTCCAGCGAAGACGCGGTAGTCCACCGCT 259  
DB 37781 CCTTGTCTGATGATCTCCGGCGGTAGTTCAGGCCCAACAGCAGGTAGTCCGGTACAT 37722  
QY 260 CGGGCGTGAACCGGTCCGGGTGCGCACCGGGATGTGTCGGGGGTGACCGGCCCT 319  
DB 37721 GGTCCGCTGCTGTTCCGGTGACCATCGGAATTTCTGACCGCTGTTTCCCT 37662  
QY 320 GCTTGGCGCGGTGCTGTCGCACACCCAGGAGACAGGTCCCGACCGATGCCGAGAGT 379  
DB 37661 GTTTCAGAGCTGTTGTCTACTACAGAACCGGATCTCTGTATGCCACCGCATGCCG 37602  
QY 380 TCCTACGCTGCGCTCTTTCGGCGTTCGCGGTACGCCACACCGCTTTGCCCTCGGCTT 439  
DB 37601 CGAGCAGTGCACCTGCTTTAGTTGTGCGCGCTATCCAGCCACTGACTTGCCTCTGCCA 37542  
QY 440 TGAAGCGAGTTACAGCGCGGACGAGTGTGCGGTGCTGCGATGCCCTCGACGTCGGCGG 497  
DB 37541 CCAGCGCATACCACTTCTCGGATCTGTGCGGAACACTGTTCCGTACGCTCGCGGAA 37484

RESULT 27  
ABZ66778/c  
ID ABZ66778 standard; DNA; 624 BP.  
XX ABZ66778;  
XX  
XX 27-OCT-2003 (revised)  
XX 21-MAR-2003 (first entry)  
XX  
XX Orthosomycin biosynthetic polynucleotide SEQ ID NO 218.  
XX  
XX Orthosomycin; biosynthesis; evernimycin; avilamycin; gene; ds.  
XX  
XX Micromonospora carbonacea; aurantiaca.  
XX  
XX WO200279505-A2.  
XX  
XX 10-OCT-2002.  
XX  
XX 28-MAR-2002; 2002WO-CA000432.  
XX  
XX 28-MAR-2001; 2001US-0279095P.  
XX 30-MAR-2001; 2001US-027909P.  
XX 20-APR-2001; 2001US-0285214P.  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Farnet CM, Zazopoulos E, Staffa A;  
XX  
XX

XX WPI: 2003-058435/05.  
DR P-PSDB; ABP99315.  
XX  
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene  
cluster, by detecting presence of nucleic acid sequence corresponding to  
PT 17 of flambamycins protein families.  
XX  
XX  
PS Claim 16; Page 368-369; 511pp; English.  
XX  
CC The invention relates to identifying orthosomycin biosynthetic genes and  
its fragment/gene cluster (AB266670-AB266681), comprising detecting the  
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
CC ABP99362). The method is useful for identifying an orthosomycin  
CC biosynthetic gene, gene fragment or gene cluster, especially an  
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
CC gene fragment or gene cluster. The method is useful for detecting the  
CC presence of any organism that contains DNA for the production of  
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
CC orthosomycins) regardless of the level at which genes for orthosomycin  
CC production are expressed by the organism or the amount of orthosomycin  
CC produced by the organism. This allows for the detection of new  
CC orthosomycin natural products, not produced by the organism. (Updated on  
CC 27-OCT-2003 to standardise OS field)  
XX  
XX Sequence 624 BP; 83 A; 225 C; 201 G; 115 T; 0 U; 0 Other;  
SQ  
Query Match 11.9%; Score 59.4; DB 7; Length 624;  
Best Local Similarity 71.6%; Pred. No. 0.032;  
Matches 78; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 1 GGTACCGGACCGGTCCGGAACACGAGTCGAGATACGCGGAGAGAACACCCCGGT 60  
DB 127 GATATCGACGCTCTGGTGAACACCGAGTCGAGTCGCGGACGAGAACCCCGCG 68  
QY 61 AGTCGGGTAGACGGTGGCGCGGAAGCGGTACCGGCTTCGACGGTCAG 109  
DB 67 GGTCCGGAAGACCCCGGGGTGAAGACGTACCGCGCCCTCGACGGCGAG 19  
RESULT 28  
AAT70152/c  
ID AAT70152 standard; DNA; 2122 BP.  
XX  
XX AAT70152;  
XX  
XX 17-OCT-2003 (revised)  
DT 23-JUL-1997 (first entry)  
XX  
DE S.longisporoflavus staurosporin synthesis gene cluster 2.1kb fragment.  
XX  
XX Staurosporin; gene cluster; indole-carbazole alkaloid; ICA; antibiotic;  
XX antiproliferative; platelet aggregation; fungus; yeast; calcium ion;  
XX Ca2+/phospholipid-dependent serine/threonine protein kinase; ds.  
XX  
XX Streptomyces longisporoflavus; (strain R19).  
XX  
XX Key Location/Qualifiers  
FH 148..768  
CDS /\*tag= a  
FT /\*label= Gene 2  
FT /note= "Encodes a protein containing 207 amino acids  
FT significantly similar to known S-adenosyl methionine-  
FT dependent methyl transferases especially those of  
FT Streptomyces and Actinomycetes"  
FT 844..1686  
CDS /\*tag= b  
FT /\*label= Gene 1  
FT /note= "Encodes a protein containing 280 amino acids  
FT significantly similar to the dUDP-4-keto-6, deoxyglucose  
FT 3,5-epimerase of Streptomyces glaucosces i.e. has 48.6%  
FT homology"  
FT  
FT

PN WO9708323-A1.  
XX  
XX  
PD 06-MAR-1997.  
XX  
PF 19-AUG-1996; 96WO-EP003643.  
XX  
XX 30-AUG-1995; 95EP-00810534.  
XX  
XX (CIBA ) CIBA GEIGY AG.  
XX  
XX Schupp T, Engel N, Bietenhader J, Toupet C, Pospiech A;  
PI WPI; 1997-179280/16.  
XX  
DR P-PSDB; AAW17977; AAW17978.  
XX  
XX Indole-carbazole alkaloid biosynthesis gene cluster - especially coding  
FT for the antibiotic staurosporin from Streptomyces longisporoflavus.  
XX  
XX Claim 7; Page 32-33; 55pp; English.  
PS  
XX The present sequence represents the 2.1kb BglII fragment of Streptomyces  
XX longisporoflavus R19 which is involved in the biosynthesis of indole-  
XX carbazole alkaloids (ICA). The sequence contains two functional fragments  
XX as indicated in the features table. The DNA or a hybrid vector containing  
XX it can be used to prepare an ICA or derivatives and precursors, either by  
XX allowing production in previously incapable organisms or by improving  
XX yields. In particular, the antibiotic staurosporin can be produced.  
XX Staurosporin is known to have inhibitory activity against fungi, yeasts,  
XX and Ca2+/phospholipid-dependent serine/threonine protein kinases (PKCs).  
XX Staurosporin also has antiproliferative activity and can inhibit platelet  
XX aggregation. The present sequence can also inactivate ICA biosynthesis  
XX genes and can be used in PCR amplification. An advantage of this is that  
XX productivity of staurosporin-synthesising Streptomyces is improved over  
XX natural strains yielding only low concentrations. (Updated on 17-OCT-2003  
XX to standardise OS field)  
XX  
XX Sequence 2122 BP; 334 A; 803 C; 670 G; 315 T; 0 U; 0 Other;  
SQ  
Query Match 11.4%; Score 56.8; DB 2; Length 2122;  
Best Local Similarity 66.1%; Pred. No. 0.087;  
Matches 82; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
QY 1 GGTACCGGACCGGTCTCCGGAACACGAGTCGAGATACGCGGAGAGAACACCCCGGT 60  
DB 271 GGGTGGCGTGGGGCTCGGTGAACGTCTTCTGTTACGGCGAGACGAACTTGCCTCTGT 212  
QY 61 AGTCGGGTAGACGGTGGCGCGGAAGCGGTACGCGCTTCGACGGTCAGCGGGCGGCGG 120  
DB 211 GGTCCGGAAGACCGCGGGGTGAACGACGCGCTCCTCGACGGTGGCGGGGTGTT 152  
QY 121 ACAC 124  
DB 151 TCAC 148  
RESULT 29  
AAC85191  
ID AAC85191 standard; DNA; 9994 BP.  
XX  
XX AAC85191;  
XX  
DT 14-MAY-2001 (first entry)  
XX  
DE S. avermitilis 10 kb genomic DNA.  
XX  
XX Glycosylation; avermectin; aglycone; biocatalysis; antibiotic;  
XX cleandrose; ds.  
XX  
XX Streptomyces avermitilis.  
OS  
XX  
XX Key Location/Qualifiers  
FH 508..1332  
CDS /\*tag= a  
FT





antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pt\_sequences

SQ Sequence 837 BP; 158 A; 290 C; 264 G; 125 T; 0 U; 0 Other;

Query Match 11.0%; Score 54.8; DB 7; Length 837;  
Best Local Similarity 46.2%; Pred. No. 0.22;  
Matches 182; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 106 TCAGCGGGGGGGACACCGCGGTACGTCGTACGTCGAGCGGGGACGTCACAGG 165  
DB 539 TCAGCTGGTGGTGGCCACGATGCCGTATTTCGGGTGTCTTCGGCGGCACCTCCTGT 480

QY 166 ATCACTGTCGCGCAGCCCGCGGAACTCCTGCTTCGCCCATGATCTGTCGCGGTGG 225  
DB 479 ACGCCAGACAGCTGCGCTTGGCGGCCACCGCCACGTCGATCAGCTGCTTGAGCACCGGC 420

QY 226 TTCAGCGAAGACGACGCGTAGTCCACCGGTGGGGGTGAACGCGTCGCGGGTGGCG 285  
DB 419 GTCTCGGCTCGATAGGTATCGGCGCAGCAGCAGCGAAGGGTCTGTGCTACCGCC 360

QY 286 ACCGGGATGTGCTGCGGGGGTGAAGCGGCGCTGCTTCGCCGGGTGTCGTCACACC 345  
DB 359 GCGCGGCGCTCAGCACGCGCATGTCCAGCGCCAGCGGGCGCGATTGGCGGATGTACAGG 300

QY 346 CAGGAGACAGGTCCGACCGATGCGCGAGAGTTCGTACCGTGGCGCTCTTCGCGCTC 405  
DB 299 CAGTTCAGTTCGCGGGGAGATGTGCGGACCGAGCCAGCAGATTCGTGCTTGCCCTTC 240

QY 406 GCGCCGTACGCCACACCGCTTGCCTTCGGCTTCAGCGAGTTCAGAGCGCGACGAG 465  
DB 239 GCTTCCAGGTGCGTTTCCAGCTCGGGGGCGGATCGAATGCTTTCGATGCGCGCTTG 180

QY 466 TCGTGGGATGCCCTCGAGCTCGCGGGGCGAACC 499  
DB 179 TTGCGGCGCGGTGACGAAGATGAGTGGTGATCC 146

RESULT 31

AAI61173/c

ID AAI61173 standard; DNA; 9975 BP.

XX AAI61173;

XX AAI61173;

22-SEP-2003 (first entry)

Actinosynnema pretiosum polyketide synthase (PKS) gene #4.

Maytansinoid; ansamitocin; antitumor; polyketide synthase; enzyme; PKS;

Gene; ds.

XX

OS Actinosynnema pretiosum.

XX WO2003045312-A2.

XX PD 05-JUN-2003.

XX PF 21-NOV-2002; 2002WO-US037547.

XX PR 21-NOV-2001; 2001US-0332158P.

XX PA (UNIW ) UNIV WASHINGTON.

XX PI Floss HG, Yu T, Leistner E;

XX DR WPI; 2003-493374/46.

XX Novel maytansinoid produced by bacterial host cell transformed with

PT expression vector comprising open reading frame from ansamitocin gene

PT cluster 1 of Actinosynnema pretiosum, useful as antitumor agent.

XX Disclosure; Page 62-67; 160pp; English.

XX The invention relates to maytansinoid produced by bacterial host cell

XX transformed with expression vector comprising open reading frame from

XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is

XX useful as a potent antitumor agent. The present sequence is A. pretiosum

XX ansamitocin gene cluster I polyketide synthase (PKS) gene

XX SQ Sequence 9975 BP; 947 A; 4102 C; 3795 G; 1131 T; 0 U; 0 Other;

Query Match 10.9%; Score 54.6; DB 7; Length 9975;

Best Local Similarity 48.7%; Pred. No. 0.19;

Matches 210; Conservative 0; Mismatches 214; Indels 7; Gaps 2;

QY 62 GTCGGGTACAGTCGGCGCGAAGCGGTACGCGCTTCGACGCTCAGCGGGCGGGCGGA 121

DB 4227 GGCAGGTGACAGCCGCGTGGCGGTGCGCGGGGACCG-CGTGAGCGCGCGG 4169

QY 122 CACCGCGGTGAGTCTGTCACTGTACGCGGGGACGTACAGGATCCATGTCGCCAG 181

DB 4168 CGAGGGCGCGCGTTCGCGACGTCGACGCGCGGCGACACCGCGCGCGCGCGCG 4109

QY 182 CCCGGCGGAATCTCTGCTCTTCGCCATGATCTCGCGGTGCTTCCAGGCGAAGCA 241

DB 4108 TCAGCTCCGCGACAGAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4049

QY 242 GCGCGTAGTCCACCGCGCTCGGGCGGTGAACGCGTCCGGGGTGCACACCGGATGTGCGTGC 301

DB 4048 GCAGTGGCG 3989

QY 302 CGGGGTGAGCG 361

DB 3988 CGGTGCGCGCGGTACAGGACGCGTCCCGGCTCCAGGACGGGTGCGCGCGCGCGCG 3929

QY 362 GACCGATGCGCGCAGAGTTCGTTCAGGTGGCGGTCTTTCGCCGTTCGCCCGCGTCCGCCACCA 421

DB 3928 GGGCGCGCGCGCGCAACCTCGCGCGGTGACCCGCGCGCGCGCGCGCGCGCG 3875

QY 422 CCCGCTGCG 481

DB 3874 GGGGCTCCGCTCG 3815

QY 482 CGAGTCGGCG 492

DB 3814 CCACGAGCG 3804

RESULT 32

AAI61224

ID AAI61224 standard; DNA; 82746 BP.

XX AAI61224;

XX AAI61224;

DT 22-SEP-2003 (first entry)  
XX Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
XX Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX Actinosynnema pretiosum.  
XX WO2003045312-A2.  
XX 05-JUN-2003.  
XX 21-NOV-2002; 2002WO-US037547.  
XX 21-NOV-2001; 2001US-0332158P.  
XX (UNIW ) UNIV WASHINGTON.  
XX Floss HG, Yu T, Leistner E;  
XX WPI; 2003-493374/46.  
XX Novel maytansinoid produced by bacterial host cell transformed with  
XX expression vector comprising open reading frame from ansamitocin gene  
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX Claim 7; Page 105-152; 160pp; English.  
XX The invention relates to maytansinoid produced by bacterial host cell  
XX transformed with expression vector comprising open reading frame from  
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
XX useful as a potent antitumour agent. The present sequence is A. pretiosum  
XX ansamitocin biosynthetic gene cluster I  
XX SQ Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;  
Query Match 10.9%; Score 54.6; DB 7; Length 82746;  
Best Local Similarity 48.7%; Pred. No. 0.16;  
Matches 210; Conservative 0; Mismatches 214; Indels 7; Gaps 2;  
QY 62 GTCCGGTAGACGCTGCGCGAAGGCTACGCGCTTCAGCTGACGGCGCGGCGGA 121  
Db 62974 GGCACAGTGCACGACGCGGTGAGCGGGTGCAGCGCGGGACCG-CGTCCGAGCGCGCG 63032  
QY 122 CACCGCGGGTCACTGCTCAGCTACGCGGGGAGCTACAGGATCCACTGTCCGCCAG 181  
Db 63033 CGAGGCGCGCGGTGCGCGACGTCGCGCGCGGACACACCGCGGCGCGCGCGCG 63092  
QY 182 CCGCGCGGAATCT 241  
Db 63093 TCAGTCCGCGACAGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 63152  
QY 242 GCGCGTAGTCCACCGCGTCCGCGGTGACGCGTCCGCGGTGCGCGGTGCGCGGTGCG 301  
Db 63153 GCAGGTGCGCGCGCGCGCGCGGTGACGAGTGGCGGGGACCGCGCGCGCGCGCGCG 63212  
QY 302 CCGCGGTGAGCG 361  
Db 63213 CGGTCCCGCGGTGACGAGGAGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCG 63272  
QY 362 GACCGATGCGCGAGAGTTGCTCAAGTGGCGCTCTTCCGCGTCCGCGCGGTGCGCGCG 421  
Db 63273 GCGCGCGCGCGCGCAACCTCGCGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 63326  
QY 422 CCGCGTTCGCGCTTTCAGCGAGTTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 481  
Db 63327 GCGGTTCGCGTCCGCGCGCGCGCGCGCGGTGATCAGGTCTCGCGCGCGCGCGCGGT 63386  
QY 482 CGACGTTCGCG 492  
Db 63387 CCACCGAGGACG 63397

RESULT 33  
AAC65299/c  
ID AAC65299 standard; CDNA; 3510 BP.  
XX  
AC AAC65299;  
XX  
DT 09-FEB-2001 (first entry)  
XX  
DE Maize Scarecrow nucleotide sequence.  
XX  
KW Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;  
XX transgenic plant; cell division; molecular marker; herbicide resistance;  
XX salt resistance; pathogen resistance; insect resistance; ss.  
XX  
OS Zea mays.  
XX  
PN WO200053723-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 07-MAR-2000; 2000WO-US005875.  
XX  
PR 10-MAR-1999; 99US-00265585.  
XX  
PA (UUNY ) UNIV NEW YORK STATE.  
XX  
PI Benfey PN, Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L;  
PI Helariutta Y, Bruce W, Lim J;  
XX WPI; 2000-594315/56.  
XX  
DR P-PSDB; AAB28595.  
XX  
XX Scarecrow gene useful for producing transgenic plants expressing genes  
XX whose product increases starch, lignin or cellulose biosynthesis and  
XX confers herbicide, pathogen or insect resistance.  
XX  
XX Disclosure; Fig 25; 200pp; English.  
XX The present sequence is given in a specification relating to the  
XX structure and function of a regulatory gene SCARECROW (SCR). SCARECROW-  
XX like (SCL) genes encoding proteins containing an amino acid sequence  
XX similar to the sequence of MOTIF III (VHII) of Arabidopsis SCR protein  
XX are disclosed. SCR and SCL genes are useful for producing transgenic  
XX plants whose cell division is modified and root and/or stem development  
XX and gravitropism of stem or hypocotyl is altered. Cell division is  
XX increased or decreased in roots resulting in thicker or thinner root  
XX development. The transgenic plants are useful for expressing a gene of  
XX interest encoding a gene product that confers herbicide, salt, pathogen  
XX or insect resistance in root or embryos and genes encoding starch, lignin  
XX or cellulose biosynthesis in shoots. The SCR gene also confers less  
XX susceptibility to lodging in the transgenic plants than a wild-type  
XX plant. SCR gene sequences are also useful as molecular markers for a  
XX quantitative trait e.g. root or gravitropism trait in molecular breeding  
XX of crop plants  
XX SQ Sequence 3510 BP; 651 A; 1225 C; 946 G; 688 T; 0 U; 0 Other;  
Query Match 10.9%; Score 54.4; DB 3; Length 3510;  
Best Local Similarity 49.0%; Pred. No. 0.23;  
Matches 145; Conservative 0; Mismatches 151; Indels 0; Gaps 0;  
QY 200 CTTTCGCATGATCTGTCGCGGTGTTCCAGCGAGAGAGCGCGGTATCCACCGCGT 259  
Db 3111 CCAGCGGTGAGCGAGAGAGTCTCTCCACCGAGCTTCAGCGCGCGCTTCTCTCCA 3052  
QY 260 CCGCGGTGAAACGCGTCCGGGGTCCGACCGGGATGTCGTGCGGGGGTGGAGCGCGCT 319  
Db 3051 CGAGCGTGTACCGTCCGAGGGGAAATCCGAGGACGAGGAGCGCTGCGCGCGCGCG 2992  
QY 320 GCTTGGCGCGGTGCTGTGTCACACCCAGAGACAGGTCCGAGACGATCCGCGAGT 379  
Db 2991 TGCCGCGAGCGAGCGCGCGGCGGACCGGACTGCGCGAGCTTCTCGCGCGAGTCCCGA 2932

QY 380 TCGTCACGGTGGCGCTCTTCGCGCTGCGCGGTACGCGACACACCGCTTGCCTCGGCCT 439  
Db 2931 ACTTACGTCGCGCGGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2872  
QY 440 TGAGCGAGTTCAGCAGGCGGACAGTCCGTCGCGATGCCCTCGACGTCGCGCGCG 495  
Db 2871 ACAGCAGCTGCTGCTCCACGAGTGCCTGCGCGCTGCTGCTGCGCGTACGCGG 2816

RESULT 34  
ADA70431  
ID ADA70431 standard; DNA; 2964 BP.  
XX  
AC ADA70431;  
DT 20-NOV-2003 (first entry)  
XX  
DE Rice gene, SEQ ID 3754.  
XX  
KW Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.  
XX  
OS Oryza sativa.  
XX  
FN WO2003000898-A1.  
XX  
PD 03-JAN-2003.  
XX  
PF 22-JUN-2001; 2001WO-IB001105.  
XX  
PR 22-JUN-2001; 2001WO-IB001105.  
XX  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tac Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX  
DR WPI; 2003-175290/17.  
XX  
PT Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
PS Claim 6; SEQ ID NO 3754; 899pp; English.  
XX  
CC The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.

QY 129 GGCTCAGCTCGTCACGTGACGCGGGGAGCGTACAGATCCACTGCCCGACCGCGCG 188  
Db 120 GGCCAGCGCGCTCGTCGCGGTCGCTGCTGCTTCCAGCTCGCTGCGCGG 179  
QY 189 GAACCTCTGCTCTTCCCATGATCTGTCGCGGTGCTTCCAGCGAAGCAGCGCGTA 248  
Db 180 GCTTGTGACAAGTACGCGCGCGCTGTCGCGGCGGCTCCAGCCGAAACAGCGGGATCA 239  
QY 249 GTCCACCGCTGCGCGGTGAACGCTCGCGGGTGCACCGGATGCTGCTGCGGGGT 308

Db 240 GCAGAGTTCACCTGSGCTGSGCCCGCGCGGAGGCGCGCGCGAGGTGTGAGTGTCTC 299  
QY 309 GAGCGCGCGCTTGTGCGCGCGCTGCTGTCGACACCCAGAGACACAGTTCGGACCGAT 368  
Db 300 GAGCGCGCGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 359  
QY 369 GCCGCGAGAGTTCGTTCACGCTGCGCTCTTTCGCGCGCTGCTGCGCGCGCGCGCGCG 428  
Db 360 GCGCGGAGAGCGCGCGCTGCTGCTGCTGCGGACGGGCTGCGCGCGCTGCTCTCAA 419  
QY 429 GCCCTGCGCTTTCAGCGAGTTCAGAGGCGGAGCAGGTGCTGCGGATGCCCTCGACGTC 488  
Db 420 GAGCGTGGACACGCGCGCTTACGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479  
QY 489 GCGCGCGGAACT 500  
Db 480 CAGCGCGCTGCT 491

RESULT 35  
ACA37987/c  
ID ACA37987 standard; DNA; 1551 BP.  
XX  
AC ACA37987;  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #19644.  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
OS Mycobacterium avium.  
XX  
FN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-023926/02.  
DR P-PSDB; ABU34117.  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 25857; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for cell  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1551 BP; 198 A; 576 G; 560 G; 217 T; 0 U; 0 Other;

Query Match 10.7%; Score 53.4; DB 7; Length 1551;  
 Best Local Similarity 47.0%; Pred. No. 0.37;  
 Matches 165; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 125 CGCGGGTACGTCGTCAGTGTACGCGGGACGTACAGATCCACTGTCGCGCAGCCC 184  
 DB 1174 CGCGGATCCGCTGCGCCGCTTCGCGGGGAGACGGCAGATGCGCAGTGGGGGGA 1115  
 QY 185 GCGGAATCTTCCTTCGTCATGATCTCGTCGGCTGTTCAGCGAAGACGAGCG 244  
 DB 1114 CGGCCACACCGCTCGTCGTCAGGTTCAGTCCCGCTGCGCGAGCAGCT 1055  
 QY 245 CGTACTCCACGCGTCGGGCTGACGGTTCGGGGTGGCACCGGATGTCGTCGGG 304  
 DB 1054 CGAGTGGCGCGGAGGTGCAATGTGTCGCGCGCCGCTGTGCGACGTGCACCGCC 995  
 QY 305 GGGTGAGCGGCGCTTCGTCGGCGGTGTCGCACACCCAGACACGAGTCCGCGAC 364  
 DB 994 CGCGGATCACCAGCGCATAGCCGCGTCTCGGGGCGTACACGTACAGGTGTGTCGG 935  
 QY 365 CGATCCGCGAGTGTGTCACGTGGGCTTCGCGCTGCGCGCTGACGCCACACCC 424  
 DB 934 AGCTGGGAGGAACCGCGCATGCCACAGCAGCTCGGCGCCCGCCATGCGTCCACGT 875  
 QY 425 GCTTCCTCGCCCTGAGCGAGTTCAGCAGGGCGAGCAGGTGCGTGCGGA 475  
 DB 874 GGGAGCCACCGACACGGGCAAGCCAGCGGTGCGGCACGACCTGGCCGA 824

## RESULT 36

AA017185/c  
 ID AA017185 standard; DNA; 27541 BP.

XX AA017185;

XX 29-NOV-2001 (first entry)

DE Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.

XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
 KW antifungal; antibiotic; nys2; ds.

XX Streptomyces noursei.

XX Key Location/Qualifiers  
 FH complement (454..1191)  
 CDS /tag= a

FT /product= "nysF protein"

FT /note= "CDS does not include start codon"

FT complement (1275..3092)

FT /tag= b

FT /product= "NysG protein"

FT CDS

complement (3070..4824)

FT /tag= c

FT /note= "CDS does not include start codon"

FT /product= "NysH protein"

FT 5122..6156

FT /tag= d

FT /product= "NysD3 protein"

FT 6338..27541

FT /tag= e

FT /product= "NysI partial protein"

FT /note= "CDS does not include stop codon"

XX WO200159126-A2.

XX 16-AUG-2001.

XX 08-FEB-2001; 2001WO-GB000509.

XX 08-FEB-2000; 2000GB-00002840.

XX 10-APR-2000; 2000GB-00008786.

XX 14-APR-2000; 2000GB-00009387.

XX (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.

XX (ALPH-) ALPHARMA AS.

XX (SINV-) SINVENT AS.

XX (DZIE/) DZIELEWSKA H.

XX (ZOTC/) ZOTCHEV S B.

XX (SEKU/) SEKUROVA O N.

XX (FJAE/) FJAERVIK E.

XX (BRAU/) BRAUTASET T.

XX (STRO/) STROM A R.

XX (VALL/) VALLA S.

XX Zetchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX WPI; 2001-557614/62.

XX P-PSDB; AA010138, AA010139, AA010140, AA010141, AA010142.

XX New nystatin polyketide synthase polynucleotides and polypeptides, useful  
 as antibiotics and antifungals.

XX Claim 2; Page 151-166; 266pp; English.

XX The present invention relates to the cloning and sequencing of the gene  
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme  
 CC involved in the biosynthesis of the macrolide antibiotic nystatin. The  
 CC nystatin PKS is useful as antifungal antibiotics. The present sequence is  
 CC a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster

XX Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 U; 0 Other;

Query Match 10.7%; Score 53.4; DB 4; Length 27541;

Best Local Similarity 47.9%; Pred. No. 0.3; Indels 6; Gaps 1;

Matches 190; Conservative 0; Mismatches 201;

QY 105 GTCAGCGGCGCGGACACCGCGGTCAGTCTGTCAGTGTACGCGGGGACGTACAG 164

DB 18955 GCGGCGGACAGCAGGAGCGGTGCGTCCGCTGTCGCGCGCAGGTCGCGCGC 18896

QY 165 GATCCACTGTCGCGCAGCGCGGAGTCTGCTCTTCGCGCATGATCTGTCGCGCTG 224

DB 18895 GACGCGCGCAGCGCGGCTGTCGATCGCTCTTCGACCGCATGACGACGACCGG 18836

QY 225 GTTCCAGGCGAAGCAGCGCGGTAGTCCACCGCGTC-----GGGCGTGAACGCGTCCG 278

DB 18835 CGAGCTGACCTCGACGATGCGGCTACTCCGCGCGCAGCAGGTCCGCCACCGTCCG 18776

QY 279 GGTGCGCAGCGGATGTGCGTCCCGGGGTGAGCCCGGCTGCTTGGCGCGCTGTCGTC 338

DB 18775 GAACCGCACCGCTCCCGCAGGTTGCGAACCAGTAGCCGCGGTCCATCCGCGCGTGC 18716



OS XX Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 363..2435  
FT /\*tag= a  
FT /product= "Human ANIC-BP-1 ligand"  
XX  
XX WO200170771-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 20-MAR-2001; 2001WO-EP003149.  
XX  
XX 21-MAR-2000; 2000EP-00106110.  
XX  
XX (MERE ) MERCK PATENT GMBH.  
XX  
XX Den Daas I, Duecker K, Hock B;  
XX  
XX WPI; 2001-607519/69.  
DR P-PSDB; AAE10857.  
XX  
XX Novel acute neuronal induced calcium binding protein type 1 ligand  
PT polypeptides useful in the treatment of stroke, head trauma, multiple  
PT sclerosis, Parkinson's disease, Alzheimer's disease and spinal cord  
PT injury.  
XX  
XX Claim 4; Page 36-39; 46pp; English.  
XX  
XX The invention relates to human acute neuronal induced calcium binding  
CC protein type 1 (ANIC-BP-1) ligand polypeptides and polynucleotides.  
CC Sequences of the invention are useful for treating human diseases  
CC including stroke, head trauma, multiple sclerosis, Parkinson's disease,  
CC Alzheimer's disease and spinal cord injury. They are also useful as  
CC vaccines. ANIC-BP-1 ligands are useful for identifying membrane bound  
CC soluble receptors. Polynucleotides of the invention are useful as  
CC diagnostic reagents, for chromosome localization studies, and as valuable  
CC tools for tissue expression studies. They are also useful in gene  
CC therapy. The present sequence is human ANIC-BP-1 ligand cDNA  
XX  
XX Sequence 2700 BP; 436 A; 901 C; 857 G; 506 T; 0 U; 0 Other;  
SQ  
Query Match 10.6%; Score 53.2; DB 4; Length 2700;  
Best Local Similarity 49.6%; Pred. No. 0.39;  
Matches 136; Conservative 0; Mismatches 138; Indels 0; Gaps 0;  
QY 218 CGCGTGTTCAGCGAAGACGCGGCTAGTCACCGGCTCGGCGTGAACGCGTCG 277  
Db 326 CGCGGCGCGCGAGCGGCTGCGGCTGCTGGGCGCAGATGGGCGCGTGCACGGGCCA 385  
QY 278 GGGTGCACCGCGGATGCGTGGCGGGGTGAGCGGCCCTGCTTGGCGGCTCGTGT 337  
Db 386 TGGTGGGTGCGGTGGCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 445  
QY 338 CGCACACCGAGACACGAGTCCGACCGATGCCGCGAGAGTTTCGTACCGTGGCGTCT 397  
Db 446 AGGACGCGCTGAGCGCGCGTCCGCGAGTGCAGCAGGCTTTCGCGCTGAAGCAGG 505  
QY 398 TCGCGTGGCGGCTAGCCACCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCG 457  
Db 506 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 565  
QY 458 CGAGCAGGTGCGGTGCGGATGCCCTCGAGTCGCGC 491  
Db 566 CTGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 599  
RESULT 39  
ACA37556  
ID ACA37556 standard; DNA; 1200 BP.  
XX  
XX ACA37556;  
XX

DT 19-JUN-2003 (first entry)  
XX  
XX Prokaryotic essential gene #19213.  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX  
XX Mycobacterium avium.  
XX  
XX WO200277193-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX  
XX 06-SEP-2001; 2001US-00948993.  
XX  
XX 25-OCT-2001; 2001US-0343923P.  
XX  
XX 08-FEB-2002; 2002US-00072851.  
XX  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU33686.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 25426; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1200 BP; 166 A; 440 C; 443 G; 151 T; 0 U; 0 Other;  
SQ  
Query Match 10.4%; Score 52.2; DB 7; Length 1200;  
Best Local Similarity 52.0%; Pred. No. 0.63;  
Matches 117; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
XX

```
QY 266 TGAACGGCTCCGGGTGCGCACCGGGATGTCGTCCCGGGGTGAGCCGCCCTGCTTGG 325
Db 422 TGCTCAGCCCGGGGTGTTGCGACGATCTCCGGCGGTGATGCTGATCCCGGGCCGG 481
QY 326 CCGGGCTCGTGTGCGACACCCAGAGAGACAGGTCCGGACCCGATGCCGAGAGTTCTGCA 385
Db 482 CCGACATCCCGCGGCCCTGCTGCGCCCTGTCGAGGCGATCGTGGCTACCGCGCA 541
QY 386 CCGTGGCGCTCTTCGCGCTGCGCGGTACGCCACACCGCTTGCCTCGCGCTTGAGG 445
Db 542 AGGAATCGACCGCGCGCTGCGCGCGCACTCGGCGGTCAACCCCTCGACGCCGCTCACCG 601
QY 446 AGTTACGAGGCGAGCAGGTGCGGATGCCCTCGACGTCGG 490
Db 602 TCGCGCAAGTGGCGGTGCGGCTGCTGCGTCAGCAACTGGCGCGG 646

RESULT 40
AAL61170/c
ID AAL61170 standard; DNA; 14055 BP.
XX
AC AAL61170;
XX
DT 22-SEP-2003 (first entry)
XX
DE Actinosynnema pretiosum polyketide synthase (PKS) gene #1.
XX
KW Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;
KW gene; ds.
XX
OS Actinosynnema pretiosum.
XX
PN WO2003045312-A2.
XX
PD 05-JUN-2003.
XX
PF 21-NOV-2002; 2002WO-US037547.
XX
PR 21-NOV-2001; 2001US-0332158P.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Floss HG, Yu T, Leistner E;
XX
DR WPI; 2003-493374/46.
XX
CC The invention relates to maytansinoid produced by bacterial host cell transformed with
CC expression vector comprising open reading frame from ansamitocin gene
CC cluster I of Actinosynnema pretiosum, useful as antitumor agent.
CC
PS Disclosure; Page 45-53; 160pp; English.
SQ Sequence 14055 BP; 1304 A; 5585 C; 5518 G; 1648 T; 0 U; 0 Other;

Query Match 10.4%; Score 52; DB 7; Length 14055;
Best Local Similarity 46.4%; Pred. No. 0.56;
Matches 206; Conservative 0; Mismatches 235; Indels 3; Gaps 1;

QY 53 CCCCCTAGTCCGGGTAGACGGTGGCGCGGAGGGGTACGCCCTTCGACGGTCAAGCG 112
Db 6787 CTTGCGGTTCCCGGTGGCGGCGAGCGGCGGACGAGGGCGCGGAGGGGGTCTGCT 6728
QY 113 GCGGCGGACACCGCGGGGTGAGTCTGTCAGTGTACGCGGGGAGTACAGGATCCAT 172
Db 6727 CCAGGTGACACCGCGGAGCGCGCGCTGCTGCGGGGTGCGCAGCAGCCCGC 6668
QY 173 GTCCGCCAGCCGCGCGAACTCCTGCTCTTCGCCATGATCTGTCGCGGTGTTCCAG 232
```

```
Db 6667 ACAGGCGCGGCGCGGGTCTCTCGCGCGCGCGCGGTGACGAGGTCTGCTCTGCT 6608
QY 233 CGAAGAGACGCGGT---AGTCCACCGCGTCCGGCGGTGAACGCGTCCGGGGTCCGACCG 289
Db 6607 GCTCGCGCAGGGCTCTCAGGGTCAAGCCGACGAGGGTTCGGGGCTGCGCCCGCGGAGGT 6548
QY 290 GGATGTGCTGTCGGGGGTGAGCCGCTTGTGCGCGCGCTGCTGTCGTCACACCCAGG 349
Db 6547 CGAGCAGCGCGCGTTCGAGCAGCGGGCGCGCGCGTTCGAGTCCGGTGTGGACACCGCA 6488
QY 350 AGACCGAGTCCGACCGAGTCCGACAGAGTTCACGTTGCGGCTTTCGCGCTCGCGC 409
Db 6487 AGTCCAGGTGACCCACTCGACCGGAAACAGCGCGCCCGAGCGGACGAGCGGTGCGCG 6428
QY 410 CGTACGCCACACCGCGCTTGCCTTCGCGCTTGAGCGAGTTTCAGCAGGGCGAGCAGGTGCG 469
Db 6427 GGGCGGGCGCGCGCGAGGGCGTTCGACCTCCAGCAGCGCGCGCTCGGGTCTGTAGG 6368
QY 470 TCGGATGCCCTTCGAGTTCGCGCG 493
Db 6367 CGCGCAGCAGTGTGTCGTCGCGCG 6344

RESULT 41
AAL61191/c
ID AAL61191 standard; DNA; 2742 BP.
XX
AC AAL61191;
XX
DT 22-SEP-2003 (first entry)
XX
DE Actinosynnema pretiosum transcriptional activator gene #1.
XX
KW Maytansinoid; ansamitocin; antitumour; transcriptional activator; gene;
KW ds.
XX
OS Actinosynnema pretiosum.
XX
PN WO2003045312-A2.
XX
PD 05-JUN-2003.
XX
PF 21-NOV-2002; 2002WO-US037547.
XX
PR 21-NOV-2001; 2001US-0332158P.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Floss HG, Yu T, Leistner E;
XX
DR WPI; 2003-493374/46.
XX
CC The invention relates to maytansinoid produced by bacterial host cell transformed with
CC expression vector comprising open reading frame from ansamitocin gene
CC cluster I of Actinosynnema pretiosum, useful as antitumor agent.
CC
PS Disclosure; Page 81-82; 160pp; English.
SQ Sequence 2742 BP; 288 A; 1096 C; 1054 G; 304 T; 0 U; 0 Other;

Query Match 10.3%; Score 51.4; DB 7; Length 2742;
Best Local Similarity 51.1%; Pred. No. 0.83;
Matches 121; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 257 GGTCCGGCGTGAACCGTTCGCGGTGCGCGGTGCGCGGTGCGCGGTGAGCCGCG 316
```





CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC on a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 2010 BP; 309 A; 711 C; 718 G; 272 T; 0 U; 0 Other;  
Query Match 10.2%; Score 51.2; DB 7; Length 2010;  
Best Local Similarity 53.3%; Pred. No. 0.92;  
Matches 130; Conservative 0; Mismatches 113; Indels 1; Gaps 1;  
QY 24 AACGATCGAGATACGGGAGAGAGAACACCCCGGCTAGTCGGGTAGACGTTGGCGCG 83  
DB 1612 AACCTGCATACCGACCGCTATACGGAACCGCGCGCGCGCGCTGCGCCCGCGCGC 1671  
QY 84 AAGCGGTACGCGCTTCGACGGTTCAGCGGTCAGCGGCGCGGACACCGCGGCTCGTCTCAC 143  
DB 1672 GAGACG-ACCACGTGGCTTCGACGCTCGGCTCGGCTCGGCTCGGCTCGGCTC 1730  
QY 144 GTATACGGGGGACGTACAGGATCCACTGTCCGCCAGCCCGCGGAACTCTGCTCTT 203  
DB 1731 GATCGGAGCGGACGCTTACCGCGCGCGGACGCTCGGCTCGGCGACATGCGCTTGGCAA 1790  
QY 204 CGCATGATCTCGTGGCGGTTCAGGCGGAGAGACAGCGGCTAGTCCACCGCGTGGG 263  
DB 1791 CGTGGCGCGTCTCGGCGTTCAGCTTCGGGATGCGGACGCTGTTCCAGGTGTCGGG 1850  
QY 264 CGTG 267  
DB 1851 CGTG 1854

RESULT 44  
AAT76906/c  
ID AAT76906 standard; DNA; 546 BP.  
XX AC AAT76906;  
XX DT 02-JUL-1998 (first entry)  
XX DE S. glaucescens acbD DNA fragment.  
XX KW Acarbose biosynthesis; acbA gene; acbB gene; acbC gene; acbD gene;  
XX acbE gene; acbF gene; enzyme; alpha-amylase inhibitor; treatment;  
XX diabetes; ss.  
XX OS Streptomyces glaucescens.  
XX DE19622783-A1.  
XX FN 11-DEC-1997.  
XX PD 07-JUN-1996; 96DE-01022783.  
XX PP

PR 07-JUN-1996; 96DE-01022783.  
XX (FARH ) HOECHST AG.  
XX PI Decker H;  
XX DR WPI; 1998-033827/04.  
XX PT Recombinant DNA molecule comprising genes for biosynthesis of acarbose -  
XX an alpha-amylase inhibitor useful in treatment of diabetes.  
XX PS Example 2; Page 7; 35pp; German.  
XX CC This sequence encodes a fragment of the Streptomyces glaucescens GLA.O  
XX acbD gene which has been generated via PCR. The acbD gene is involved in  
XX the acarbose biosynthesis pathway. This gene and other acarbose  
XX biosynthesis genes such as those represented in AAT76903 are useful for  
XX producing acarbose, which is an alpha -amylase inhibitor useful in the  
XX treatment of diabetes

XX SQ Sequence 546 BP; 69 A; 235 C; 168 G; 74 T; 0 U; 0 Other;  
Query Match 10.1%; Score 50.6; DB 2; Length 546;  
Best Local Similarity 49.8%; Pred. No. 1.3;  
Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;  
QY 239 GCACGCGTATGTCACCGCGTTCGGGCTGACCGCTCCGGGTGCGCACCGGATGTGCG 298  
DB 383 GGAGGGAGCGGTATACCTCGTTCGGTGGAGACGTGACGAGGGTGGCGACACCGTGGCGGA 324  
QY 299 TGCCGCGGCTGAGCGGCGCTGCTTGGCGCGGCTCGTGTGCGCACACCCAGGACCAAGGT 358  
DB 323 CGCGGCGCTCGAGCAGGACCTGGGTGCCAGCACGTTGGTGGCGGTGAAGGCGCGCTGT 264  
QY 359 CCGGACCGATGCGCGCAGAAAGTTGTCACGGTGGCGCTCTTCGCGCTCGCGCGTACGCCA 418  
DB 263 CGGTGATGAGCGGTGCGACGTGCGACTCGGCCCGGAAAGTGCAGATGCTGTCGCGCG 204  
QY 419 CCACCGCTTCCCTCCGCTTCAGCGAGTTTCAGAGGCGGCGAGCAGTCCGTCCGGATGC 478  
DB 203 CGGCGAGGCTGTCAGCAGCGGCTGTCGCGACCGTCCCTCGCACCGTCCCTGACGAGGTGAGGCGCG 144  
QY 479 CCTCGAGTCCGCGCGG 495  
DB 143 GATGTCACGCAACGCG 127

RESULT 45  
ADA70563/c  
ID ADA70563 standard; DNA; 2697 BP.  
XX AC ADA70563;  
XX DT 20-NOV-2003 (first entry)  
XX DE Rice gene, SEQ ID 3886.  
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;  
XX gene; ds.  
XX OS Oryza sativa.  
XX FN WO2003000898-A1.  
XX PD 03-JAN-2003.  
XX PF 22-JUN-2001; 2001WO-IB001105.  
XX PR 22-JUN-2001; 2001WO-IB001105.  
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
XX  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
XX pathogenic infection for conferring resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection by determining or detecting plant  
XX gene expression.  
XX  
XX Claim 6; SEQ ID NO 3886; 899pp; English.  
XX  
XX The present invention relates to a method (M1) for identifying genes  
XX involved in plant resistance or response to pathogenic infection. M1  
XX comprises identifying a gene whose expression is significantly altered in  
XX the incompatible interaction of plant gene expression relative to  
XX expression of the gene in an uninfected plant, in a mutant plant that  
XX does not express a gene associated with response to pathogenic infection,  
XX or in a corresponding incompatible or compatible interaction. (M1) is  
XX useful for conferring resistance to resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection. The present sequence was used to  
XX illustrate the invention.  
XX  
XX Sequence 2697 BP; 446 A; 903 C; 922 G; 426 T; 0 U; 0 Other;  
XX  
XX Query Match 10.1%; Score 50.6; DB 7; Length 2697;  
XX Best Local Similarity 48.0%; Pred. No. 1.2; Indels 6; Gaps 1;  
XX Matches 180; Conservative 0; Mismatches 189; Indels 6; Gaps 1;  
XX  
XX QY 118 CGACACCGCGGGTCACTCGTCACGTCGTACGCGGGGAGTCACAGATCCACTGTCG 177  
XX Db 2528 CGCACACCGTGGGCGATCTCGGCGCGAGCGAGGCTCTCGTGAAGCAGCAGAGCGCG 2469  
XX QY 178 CCAGCCCGCGGAACCTCGTCTCTCCATGATCTCGTGGCGGTGTTCCAGCGCAAG 237  
XX Db 2468 ACGCGGACGGTCGTGCGCTTGGCCCTCGTCGACGCGCCCTCCAGCTCGCTCGACG 2409  
XX QY 238 AGCAGCGGTAGTCACCGCGTCGCGCGCTGAACGCTCCGGGTCGCGACCGGATGTC 297  
XX Db 2408 AGGTCAGGTATCTCCGCTGCTCATGGAGCTCCAGTCCAGCAGTACTCTGTCG 2349  
XX QY 238 GTGCGCGGGGTAGCGCGCCCTCTGTGCGCGGCTGCTGTCGACACCCAGGACGAG 357  
XX Db 2348 CGCGCGCGCTGTGTCCTCACTCGCGCTGACGTCATCGCGCTGCTCCATCGGACGGCAAG 2289  
XX QY 358 TCGGACCGATGCCAGAAATTCGTACGCTGGCGCTCTTCGCGCTGCGCGCTAGCGCC 417  
XX Db 2288 TCCGAGTCTCGCGCTCGCGGACGACGCGGAGCTGTTGGCGCGGCGC-----CGCC 2235  
XX QY 418 ACCACCGCTGCGCTCGCGCTTTCGAGCGAGTTACAGCGGCGAGCAGTCTGTCGATG 477  
XX Db 2234 GCTCTGTCGCGCGGTTCTTCTCCCGTGTATCATGCTGAGCAGCAGCAGCAGTCCGAA 2175  
XX QY 478 CCTCGACGTGCGCG 492  
XX Db 2174 CTGTAGAGTCTGCGG 2160  
XX  
XX RESULT 46  
XX AAL40781/c  
XX ID AAL40781 standard; DNA; 88421 BP.  
XX  
XX AC AAL40781;  
XX  
XX 03-OCT-2002 (first entry)  
XX  
XX 88421nt genomic DNA of ramoplanin producing Actinoplanes sp.  
XX  
XX Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;  
XX biosynthesis gene cluster; bioengineering; peptide synthetase module;  
XX adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;  
XX chlorinate; lipopeptide; gene; ds.  
XX  
XX Actinoplanes sp.

XX Key Location/Qualifiers  
XX CDS 2077..3078  
XX /tag= a  
XX /product= "Protein of ORF 1"  
XX 3118..4032  
XX /tag= b  
XX /product= "Protein of ORF 2"  
XX 4038..5048  
XX /tag= c  
XX /product= "Protein of ORF 3"  
XX complement(6665..5814)  
XX /tag= d  
XX /product= "Protein of ORF 4"  
XX complement(7703..6693)  
XX /tag= e  
XX /product= "Protein of ORF 5"  
XX complement(9464..8130)  
XX /tag= f  
XX /product= "Protein of ORF 6"  
XX 9691..10761  
XX /tag= g  
XX /product= "Protein of ORF 7"  
XX complement(12751..10829)  
XX /tag= h  
XX /product= "Protein of ORF 8"  
XX complement(13617..12802)  
XX /tag= i  
XX /product= "Protein of ORF 9"  
XX complement(15203..13614)  
XX /tag= j  
XX /product= "Protein of ORF 10"  
XX 15591..15863  
XX /tag= k  
XX /product= "Protein of ORF 11"  
XX 15880..19035  
XX /tag= l  
XX /product= "Protein of ORF 12"  
XX 19032..39713  
XX /tag= m  
XX /product= "Protein of ORF 13"  
XX 39713..65800  
XX /tag= n  
XX /product= "Protein of ORF 14"  
XX 65826..66530  
XX /tag= o  
XX /product= "Protein of ORF 15"  
XX 66546..67370  
XX /tag= p  
XX /product= "Protein of ORF 16"  
XX 67384..70059  
XX /tag= q  
XX /product= "Protein of ORF 17"  
XX 70099..70662  
XX /tag= r  
XX /product= "Protein of ORF 18"  
XX 70659..71906  
XX /tag= s  
XX /product= "Protein of ORF 19"  
XX complement(73439..71964)  
XX /tag= t  
XX /product= "Protein of ORF 20"  
XX complement(74216..73563)  
XX /tag= u  
XX /product= "Protein of ORF 21"  
XX complement(75424..74213)  
XX /tag= v  
XX /product= "Protein of ORF 22"  
XX 75535..76464  
XX /tag= w  
XX /product= "Protein of ORF 23"  
XX complement(78110..76449)  
XX /tag= x

FT CDS /product= "Protein of ORF 24"  
FT complement(79864..78107)  
FT /\*tag= y  
FT /product= "Protein of ORF 25"  
FT complement(81624..79861)  
FT /\*tag= z  
FT /product= "Protein of ORF 26"  
FT complement(81909..81682)  
FT /\*tag= aa  
FT /product= "Protein of ORF 27"  
FT complement(82346..82062)  
FT /\*tag= ab  
FT /product= "Protein of ORF 28"  
FT 82587..84446  
FT /\*tag= ac  
FT /product= "Protein of ORF 29"  
FT 84481..85548  
FT /\*tag= ad  
FT /product= "Protein of ORF 30"  
FT 85556..86845  
FT /\*tag= ae  
FT /product= "Protein of ORF 31"  
FT 87372..86803  
FT /\*tag= af  
FT /product= "Protein of ORF 32"  
FT 87494..88420  
FT /\*tag= ag  
FT /product= "Protein of ORF 33"  
FT  
XX WO200231155-A2.

18-APR-2002.

15-OCT-2001; 2001WO-CA001462.

13-OCT-2000; 2000US-0239924P.

12-APR-2001; 2001US-0283296P.

24-JUL-2001; 2001US-00910813.

(ECOP-) ECOPIA BIOSCIENCES INC.

Farnet CM, Zazopoulos E, Staffa A;

WPI: 2002-435445/46.

DR P-PSDB: AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,  
DR AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,  
DR AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,  
DR AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,  
DR AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.

XX Novel isolated ramoplanin biosynthetic pathway polypeptide useful for  
PT chemically modifying biological molecule that is a substrate for a  
PT polypeptide encoded by a ramoplanin biosynthesis gene cluster.

PS Disclosure; Page 87-135; 212pp; English.

XX The invention relates to an isolated ramoplanin biosynthetic pathway  
XX polypeptide selected from a polypeptide of open reading frames (ORF) 1-  
XX 32. The isolated polypeptides are useful for chemically modifying a  
XX biological molecule that is a substrate for a polypeptide encoded by a  
XX ramoplanin biosynthesis gene cluster, by contacting the biological  
XX molecule with the isolated polypeptide, where the polypeptide chemically  
XX modifies the biological molecule. The method comprises contacting the  
XX biological molecule with at least two different polypeptides encoded by  
XX ramoplanin ORFs 1-31. The polypeptides are useful for directing the  
XX biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated  
XX gene cluster comprising the ORFs is useful as a substrate for  
XX bioengineering of antibiotic structures. An isolated polypeptide or its  
XX encoding nucleic acid sequence is useful for generating derivatives of  
XX ramoplanin, for improving production or for producing variants of other  
XX antibiotics of the peptide class. The isolated polypeptides are useful  
XX for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain  
XX in conjunction with other peptide synthetase modules and allowing the

CC incorporation of Thr into a peptide antibiotic precursor, for modifying  
CC fatty acid structure and/or enhancing fatty acid incorporation into the  
CC peptide antibiotic structure, for production of an hydroxyphenylglycine  
CC (HPG)-containing peptide antibiotic, for enhancing secretion of  
CC ramoplanin or its variants and derivatives, for enhancing uptake of  
CC precursors for ramoplanin biosynthesis, for enhancing production of  
CC ramoplanin products or its variants or derivatives, to chlorinate HPG of  
CC a peptide antibiotic precursor, and for designing specific nucleotide  
CC probes and primers for identifying and isolating putative lipopeptide  
CC -producing microorganisms. This polynucleotide sequence represents the  
CC 88421nt genomic DNA of a ramoplanin producing Actinoplanes sp.  
XX microorganism of the invention

SQ Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 U; 0 Other;

Query Match 10.1%; Score 50.6; DB 6; Length 88421;

Best Local Similarity 45.3%; Pred. No. 0.87;

Matches 224; Conservative 0; Mismatches 269; Indels 2; Gaps 1;

Qy 6 CCGACCGTGTCCCGGAAACAGAGTCAGATACGGGAGAGGAACACCCCGGTAGTCC 65  
Db 71491 CCGAACGTGCCAGGCGAGCGGTTTCACCGCGGCATGTGCAGGCGGAGCCGTCAGC 71432  
Qy 66 GGGTAGACGTGGGCGGAGGCGTACGGCCCTTCACCGGTACGGGGGGGGGACACC 125  
Db 71431 GCGAACAGAGACCGCAACTCGGCCCGCGCGTGAACCGCCAGAACTCCATC 71372  
Qy 126 GGGGGGTACGTCTGTCACGTGTACGGGG--GGACGTACAGGATCCACTGTCCGCCAGCC 183  
Db 71371 CAGCGTCCGGTTGCAGGCGGACACCGGCGAGTTGACAGAGCCAGGCCCGCGGCA 71312  
Qy 184 CCGGGGAATCTCTGCTCTTCGCGCATGATCTGTGGGGTGTTCAGGGGAAGAGCAGC 243  
Db 71311 CCGCGCGCCACCGGGGGAACCCCGCATCCCGCGTGGCGGAGGAGCAGCAGCAGC 71252  
Qy 244 GCGTAGTCCACCGTCGGCGGTGAACCGCTCCGGGTGCGCACCGGATGTCCGTGCGG 303  
Db 71251 GGGCCGAGCAGCAGCAGCGGAGACAGTTGGCCCGCGTCCCGAGCCGATCAGCAGCG 71192  
Qy 304 GGGGTGAGCCGCGCTGTGTCGGCGGCGTGTGTCGCAACCCAGAGAGACAGGTCCGGA 363  
Db 71191 GCCAGCAGCGCGGCGTCCCGTTCCAGGCGAGATCGCGAGCACCGCCCGCGGAGCGCG 71132  
Qy 364 CCGATGCCGAGAGTTTCGTACGGTGGCGCTCTTCGCGTCCGCGGTACGCCACCCACC 423  
Db 71131 AGCAGTCCAGTTCACCGTCCCGGCGAGATCAGCACCGGTGCGGGCGGAGCAGCAGC 71072  
Qy 424 CGTTGCGCTCGGCGCTTGAAGCGAGTTACAGAGGCGGAGCAGGTCCGTTGCGGATGCC 483  
Db 71071 GCGTCGCGCGCGCGCGCGTCCCGTCCAGGCGAGCAGCAGCGTCCGCCACACGTCAG 71012  
Qy 484 ACCTCGCGCGGCAAC 498  
Db 71011 ACCCGAGGCGGAGC 70997

RESULT 47

AAQ10190/c

ID AAQ10190 standard; DNA; 23666 BP.

XX AAQ10190;

AC AAQ10190;

XX 25-MAR-2003 (revised)

DT 27-MAR-1991 (first entry)

DE Cephalosporin antibiotic biosynthetic genes.

XX cephalosporin; antibiotic; S-(L-alpha-aminoadipyl)-L-cysteiny]-D-;  
KW valine synthetase; isopenicillin N synthetase; isopenicillin N epimerase;  
KW deacetoxycephalosporin C synthetase; beta-lactamase;  
KW deacetoxycephalosporin C hydroxylase; ss.

OS Lysobacter lactamgenus.



CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1248 BP; 157 A; 451 C; 474 G; 166 T; 0 U; 0 Other;

Query Match 10.0%; Score 50.2; DB 7; Length 1248;  
 Best Local Similarity 50.7%; Pred. No. 1.5;  
 Matches 180; Conservative 0; Mismatches 163; Indels 12; Gaps 2;

QY 158 CGTACAGGATCCACTGTCGCGACGCGCGGAACTCCCTGCTTCGCGATGATCGT 217  
 Db 616 CGCGGACGATCACTCGCGCTCGCGGCGAGCTACCGAATCGTTGCGCCAGAGCGCG 557  
 QY 218 CGGCGGTGTTCCAGGCGAAGAGACGCGGTAGTCCACCGCGTCGGCGGTGAACGGGTCCG 277  
 Db 556 CGGCGGCCACCCACACACGAGTGGGACTGTGCTAGGCGCGGAGGTATCAGCAGTGGT 497  
 QY 278 GGTGCGCACCGGAGTGCCTGCCG-----GGGTAGCGCGCCCTGTTGGCGCGCG 331  
 Db 496 CGAGTTGGCCACGACACCGGCTCGGGTCCGGGCGAGGCGCGCGCGAGGTCCGCT 437  
 QY 332 TCCTGTGCGCACACCCAGGAGACAGGTCCGGACCGATGCCGAGAGTTTCGTACGCGTGG 391  
 Db 436 CGATGACGTTGATCTCGCGACCCGCTCAGCTGTGTCGGGACAGGTTGAGCAGCACA 377  
 QY 392 CGTCTTCGCGTCCGCGCGCTACCGAC-----CACCGCTTGCCTCGGCTTACGG 445  
 Db 376 CGACGCCCGGCGAGCGGCGTGGACACGTCGGGACGTCGATCTCGTCACTCCAGGG 317  
 QY 446 AGTTCAGCAGGCGAGCAGGTGCGGTGCGGATGCCCTCGACGTGCGCGCGGAACCT 500  
 Db 316 CGGCCAGCGGCGCGCGGTTGGCGGCGAGCGCGCGACCGAGCGCGCGGTCCAT 262

RESULT 49  
 AAX02914/c  
 ID AAX02914 standard; DNA; 1605 BP.

XX AAX02914;

XX 20-MAY-1999 (first entry)

XX Z. mays cck1 DNA coding region.

XX Cytokinin oxidase; cck1; transgenic plant; altered growth behaviour;  
 KW cytokinin-associated pathogenesis; resistance; fungi; nematode; assay;  
 KW grain yield; secondary growth; metabolism; senescence; ss.

XX Zea mays.

XX WO9906571-A1.

XX 11-FEB-1999.

XX 30-JUL-1998; 98WO-US015844.  
 PF 30-JUL-1997; 97US-0054269P.  
 PR 29-JUL-1998; 98US-00124541.  
 XX (UMOR ) UNIV MISSOURI.

XX Morris RO;

XX WPI; 1999-153800/13.

XX P-PSDB; AAW93007.

XX New cytokinin oxidase from maize - used to generate transgenic plants

PT with, e.g. better disease resistance and growth characteristics.

XX Claim 7b; Page 64-69; 140pp; English.

XX This sequence encodes a Zea mays cytokinin oxidase, cck1. This gene is  
 CC used to generate transgenic plants in which cytokinin-associated  
 CC pathogenesis or growth behaviour is altered. Particularly applications  
 CC include generation of plants with increased resistance to fungi and  
 CC nematodes, increased grain yield and superior secondary growth  
 CC properties. Host cells are used for production of recombinant cck1  
 CC nucleic acid which is useful in an assay for determining cytokinin  
 CC concentrations, and for its studying effects on plant growth and  
 CC metabolism, including senescence

XX Sequence 1605 BP; 236 A; 560 C; 561 G; 248 T; 0 U; 0 Other;

Query Match 10.0%; Score 50.2; DB 2; Length 1605;  
 Best Local Similarity 51.0%; Pred. No. 1.4;  
 Matches 171; Conservative 0; Mismatches 158; Indels 6; Gaps 2;

QY 165 GATCCATGTGCGCAGCGCGGAGAACTCTCTCTCTTCCGCTATGATCTCGTGGGCTG 224  
 Db 1194 GAGCAGCGGTGCGGACCGCGCCACAGCCCGCTTGTGAGCGCCACCTCTCGCGGTG 1135  
 QY 225 GTTCCAG3CGAAGAGCAGCGGTAGTCCACCGGTGCGGGGTGAACGCGTCCGGGTGCG 284  
 Db 1134 CACCGGTCAAGAACCGCGGTAGGCA---CGTGGGTGGAACGCGAAC---CCCTC 1081  
 QY 285 CACCGGATGTGCTGCGGGGTGAGCGCGGCGCTTGTGGCGGCGTCTGTGCGACAC 344  
 Db 1080 CACGTAGTTCAGCTGCGCCAGCAGGACGCGGAGTCTCTGTCCACCGCGCGCGCGC 1021  
 QY 345 CCAGGAGACCGGTCCGACCGATGCCGAGAGTTGTCACGCTGCGCGTCTTCGCGCT 404  
 Db 1020 CGTGGCGTGTGCTGAGTGTGAGCGTGGCTCTGATCTGTACCGTGTGGGTTCGCTC 961  
 QY 405 CGCGCGGTACGCCACCGCGCTTGCCTCGGCTTGAAGGAGTTACAGAGGCGAGCAG 464  
 Db 960 CCGCGGAGCGGACGATCCGGGCGAGCTCGGCGTGGTGAAGAACCCCGTGTTCGCGAG 901  
 QY 465 GTCGTTGCGGATGCCCTCGAGCTCGCGCGGCGAAC 499  
 Db 900 GTCGTTGCGGATGCCCTCGAGTTCAGAACACCGGACC 866

RESULT 50

AAX02913/c

ID AAX02913 standard; DNA; 6733 BP.

XX AAX02913;

XX 20-MAY-1999 (first entry)

XX Z. mays cck1 DNA.

XX Cytokinin oxidase; cck1; transgenic plant; altered growth behaviour;  
 KW cytokinin-associated pathogenesis; resistance; fungi; nematode; assay;  
 KW grain yield; secondary growth; metabolism; senescence; ss.

2928 CGTGGCGTTGCTAGTTGAGCGTGGCCCTCGATGCTGTACACGGTGGTGGGTTCCGGCTC 2869  
405 CGCGCGGTACGCCACCCCGCTTGGCTCGCCCTTGAGCGAGTTTCAGCAGGCGGAGCAG 464  
2868 CCGCGGAGCGCGACGATCCGGGCGACGTCGGCGGACGTCGGCGTGAAGAACCCCGCTTCCGCCAG 2809  
465 GTCGGTGGGATGCCCTCGACGTCGGCGGCGAACC 499  
2808 GTCCGTCGCCAGGCTCTGTTACGAAACACCGACC 2774

Search completed: June 27, 2004, 18:18:33  
Job time : 332.98 secs

OS Zea mays.  
XX Key Location/Qualifiers  
FH CDS 1497..3607  
FT /\*tag= a  
FT /product= "ckx1"  
FT 1497..2111  
FT /\*tag= b  
FT /number= 1  
FT 2112..2523  
FT /\*tag= c  
FT /number= 1  
FT 2524..3216  
FT /\*tag= d  
FT /number= 2  
FT 3217..3310  
FT /\*tag= e  
FT /number= 2  
FT 3311..3607  
FT /\*tag= f  
FT /number= 3  
XX W09906571-A1.  
FN 11-FEB-1999.  
XX 30-JUL-1998; 98WO-US015844.  
XX 30-JUL-1997; 97US-0054268P.  
PR 29-JUL-1998; 98US-00124541.  
XX (UMOR ) UNIV MISSOURI.  
XX Morris RO;  
XX WPI; 1999-153800/13.  
DR P-PSDB; AAW93007.  
XX  
XX New cytokinin oxidase from maize - used to generate transgenic plants  
XX with, e.g. better disease resistance and growth characteristics.  
XX Claim 7a; Page 54-63; 140pp; English.  
XX  
XX This sequence encodes a Zea mays cytokinin oxidase, ckx1. This gene is  
XX used to generate transgenic plants in which cytokinin-associated  
XX pathogenesis or growth behaviour is altered. Particularly applications  
XX include generation of plants with increased resistance to fungi and  
XX nematodes, increased grain yield and superior secondary growth  
XX properties. Host cells are used for production of recombinant ckx1  
XX nucleic acid which is useful in an assay for determining cytokinin  
XX concentrations, and for its studying effects on plant growth and  
XX metabolism, including senescence  
XX  
XX Sequence 6733 BP; 1786 A; 1585 C; 1592 G; 1765 T; 0 U; 5 Other;  
SQ  
Query Match 10.0%; Score 50.2; DB 2; Length 6733;  
Best Local Similarity 51.0%; Pred. No. 1.3; Indels 6; Gaps 2;  
Matches 171; Conservative 0; Mismatches 158;  
QY 165 GATCCACTGTCGCCAGCCGCGGAACTCCTCTTCGCGCATGATCTCGTCGCGTG 224  
DB 3102 GAGCCACGGGTGCGGCACCGCCACAGCCCGCAGCTGTTGAGCGCCACCTCTCGCGTG 3043  
QY 225 GTTCCAGGGGAGAGCAGCGGTAGTCACCGGTCGCGCGTGAACGGTCGCGGTGCG 284  
DB 3042 CACCCGGTCAAGGACCGCGTAGGCCA---CGTCGCGTGGAGCGGAACCC---CTC 2989  
QY 285 CACCGGATGTGGTGGCGGGGTGAGCCGGCCCTCTTGGCCGGGTGCTGTGCGCACAC 344  
DB 2988 CAGTAGCTCAGGTGCGCCAGCAGCAGCGAGCTCTGTCACCGCGCGCGCGCGC 2929  
QY 345 CCAGGAGACAGGTCCGGACCGATGCGCAGAAAGTTTCGTACGGTGGCGCTCTTCGCCGT 404

SQTGLVRSRRTDLSLAATPPAAKSSCRVTAASEGGGAAAGAAAGAPVSVAGSPLS  
PGVAPRLSRHTLSGGCGRGSKVALLPAGALQHSRSMMPVAHSPPAATPGSGLS  
SSSHGSGSPYPPGPHPLPLPHLHGPGQRPSSGASASGSPDPMGLDVGSP  
GDLRAFCHSRNTPESTAEPTAPDGGGGEFYCYMTMDRELSCGSRVSVGDAQ  
DLRLGRKRTYSLTTPAQRPVPPSSASLDVYTLMAATPSGAGRLCPGCPASSXV  
ALHPYEDYDIEIGSHRSSNUGADDGYMPTFGAALAGSGSGCKDDYPMGFA  
SVAPKQTLQPRAAAATAAAPSAGPAPATSAAGRTFPASGGYKASPSAESBED  
SGYRMWCGSLXSEHADKLLPNGLYVSPDAVTGTPDFFSAALHFGGELRG  
VPGCYSLPSYKAPYTCGSDQYVLMSPVGRILIEERLEPQATPGSQAASAF  
AGTQPPHVPVSVPRSGRPEGLGQGRVAPTRLISLEGLPSLPSMHEYLPLPAP  
KSPGEYINDFEGECARLSPPALLASASSSSLLSASSPASSLSGSGTPTSSDSRQ  
RPSLDYMLDFSPKPKGAPSGHPVGLDGLLSPEASSPYPLPSPASSPSGL  
QPPPPPPAGLYRLPPASAVATAGGGAASSLSDDGNDGYTEMAFGVAATPPQPI  
IAAPPKEARVASPTSGVRLSMEQVSGVEAFLOASQPPDHRGAKVIRADPGGGR  
RHSETFSTTTTVPSPFAHNPXHNASVENYSLRKSSEGGVGVGGGDEPPT  
SPROLQAPPLAQGRPWTPGQGLVPGCGSPMRRTSAGFQNGLYNLIADVRE  
PGLPPOPPPPPLPQPGDKSWGRTRSLGLLSAVGVSTGGCGGCGPGCALPPANT  
YASIDFLSHLKEATIVK"  
223..225  
variation  
/gene="IRS2"  
/note="compared to GenBank Accession Number AF322114"  
/replace=""

ORIGIN  
Query Match 10.5%; Score 54.8; DB 9; Length 4286;  
Best Local Similarity 46.2%; Pred. No. 41;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
QY 46 ACTGCCACTCGCGGGGTGACACAGTTCGACAGTGGAGTTGAGCAGCGCTCATCG 105  
DB 1037 AGTTCGGCGCGCGAGTAAGACCAATCGTCGGGTCGTCGCCACGACCCCATCAGCG 1096  
QY 106 CGCGCTGGCGCGATCGCGAACACCGCGCCAGGATCAGTCGACGACGCGATGCGCG 165  
DB 1097 TCCCGCGCGCGCGCGCCACACACCTGCTCACTGCTCCCGCCACGACGCGCGCTGG 1156  
QY 166 CCTCGATCTCGAGCTGCGCGCGAGTCTGATCTGCTGCGCGGCTTCCACACCGAGGAA 225  
DB 1157 TCGCGCGCTCGCGACCGACAGCTGCGCGCCACCGCGCGCGCGCGCAAGTGCAGCTCGT 1216  
QY 226 ACTGCTCGTGGCAGCGCGCGCGCGCGCGAGTCTGCTCAGGCTGGGTCACGGA 285  
DB 1217 GCGCGGTGCGACCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1276  
QY 286 CTTCTCGCGGTGCGGAGCGGAAACGCGCAGCGCTCAGATCCCTCTCAGTCGCACTCGCT 345  
DB 1277 GCGCCAGCGCGGTGTCGCTGGTGGAGC-CCCTGAGCGCGCGCGCGCGCGCGCG 1335  
QY 346 CAGTGGCGGTGCTCCCGCTTGGCTGGAGGATAGCGTTACGACGAGCGCGCACCA 405  
DB 1336 CTGAGCGCTCGCACACCTGAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCTG 1395  
QY 406 CGCGCGCGCGCGCGCGCGGTTCAGCGATCCGCTGATGACGCGGCTGCGGGTTCGGG 465  
DB 1396 CTGCGCGCGAGGCGCGCGCTGCAACACAGCGCTCAATGCTATGCCCTGCGCGACTCG 1455  
QY 466 GCGGTGCGGAGATCGTACCGCGCGGACCGCTGCGCGCGCGCGCGCGCGCGCGCG 515  
DB 1456 CCGCGCGCGCGCGCGCGCGCGCTCCCTGCTGCTGCTGCGCGCGCGCGCGCGCGCG 1505

RESULT 24  
AF385932 4289 bp DNA linear PRI 25-JUN-2001  
LOCUS Homo sapiens insulin receptor substrate 2 insertion mutant (IRS2)  
DEFINITION gene, partial cds.  
ACCESSION AF385932  
VERSION AF385932.1 GI:14537853  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 4289)  
Heyne, B.  
Two insertions in insulin receptor substrate 2 (N28\_H29insN);  
N28\_H29insNN)  
Unpublished  
JOURNAL  
REFERENCE 2 (bases 1 to 4289)  
Heyne, B., Gehrisch, S. and Jaross, W.  
Direct Submission  
TITLE Submitted (25-MAY-2001) Institut fuer Klinische Chemie und  
Laboratoriumsmedizin, Universitaetsklinikum der Technischen  
Universitaet Dresden, Fetscherstr. 74, Dresden 01307, Germany  
FEATURES  
Location/Qualifiers  
1..4289  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosomes="13"  
/map="13q34"  
/note="isolated from a patient with type 2 diabetes  
mellitus"  
<139..>4156  
/gene="IRS2"  
/db\_xref="GI:14537854"  
/product="insulin receptor substrate 2 insertion mutant"  
139..>4156  
/gene="IRS2"  
/note="N28\_H29insNN"  
/codon\_start=1  
/product="insulin receptor substrate 2 insertion mutant"  
/protein\_id="AAK66750.1"  
/db\_xref="GI:14537854"  
/translation="MASPPRHGGPPGASGDGPNLNNNNNNNNHVRKCYLRKQKH  
GKRRFVLPAGAGDEATAGGSAPOPPRLVEYSEKKWRKAGAPKRVIALDCCIN  
INKRADAKHYKLIATYKDEYFAVAENEOEGWYRALTDLVSEGRAAGDAPAAA  
PAASCSASLPALGCSAGAGAEADSYLVAPATAAYREYQVNLKPKGQSKNLTVG  
YRUCLSARTIGFVKNCEQPSVTLQMLNRRCHSDSFFIEVGRSAVTGPELWMA  
DDVSAQNTHTILEAMKALKELFEPRKRSQSGSSATFISVPGARRHHLNLP  
PSGTGLVRSRRTDLSLAATPPAAKSSCRVTAASEGGGAAAGAAAGAPVSVAGSPL  
SPGVAPRLSRHTLSGGCGRGSKVALLPAGALQHSRSMMPVAHSPPAATPGSGL  
SSSSGHGSGSPYPPGPHPLPLPHLHGPGQRPSSGASASGSPDPMGLDVGSS  
GDLRAFCHSRNTPESTAEPTAPDGGGGEFYCYMTMDRELSCGSRVSVGDAQ  
QDLRLGRKRTYSLTTPAQRPVPPSSASLDVYTLMAATPSGAGRLCPGCPASSXV  
VATHPYEDYDIEIGSHRSSNUGADDGYMPTFGAALAGSGSGCKDDYPMGFA  
SVAPKQTLQPRAAAATAAAPSAGPAPATSAAGRTFPASGGYKASPSAESBED  
ASVAPKQTLQPRAAAATAAAPSAGPAPATSAAGRTFPASGGYKASPSAESBED  
DSYMRWCGSLXSEHADKLLPNGLYVSPDAVTGTPDFFSAALHFGGELRG  
GAGTQPPHVPVSVPRSGRPEGLGQGRVAPTRLISLEGLPSLPSMHEYLPLPAP  
KSPGEYINDFEGECARLSPPALLASASSSSLLSASSPASSLSGSGTPTSSDSRQ  
RPSLDYMLDFSPKPKGAPSGHPVGLDGLLSPEASSPYPLPSPASSPSGL  
QPPPPPPAGLYRLPPASAVATAGGGAASSLSDDGNDGYTEMAFGVAATPPQPI  
IAAPPKEARVASPTSGVRLSMEQVSGVEAFLOASQPPDHRGAKVIRADPGGGR  
RHSETFSTTTTVPSPFAHNPXHNASVENYSLRKSSEGGVGVGGGDEPPT  
SPROLQAPPLAQGRPWTPGQGLVPGCGSPMRRTSAGFQNGLYNLIADVRE  
PGLPPOPPPPPLPQPGDKSWGRTRSLGLLSAVGVSTGGCGGCGPGCALPPAN  
TVASIDFLSHLKEATIVK"  
223..228  
/gene="IRS2"  
/note="compared to GenBank Accession Number AF322114"  
/replace=""

variation

ORIGIN  
Query Match 10.5%; Score 54.8; DB 9; Length 4289;  
Best Local Similarity 46.2%; Pred. No. 41;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
QY 46 ACTGCCACTCGCGGGGTGACACAGTTCGACAGTGGAGTTGAGCAGCGCTCATCG 105  
DB 1040 AGTTCGGCGCGCGAGTAAGACCAATCGTCGGGTCGTCGCCACGACCCCATCAGCG 1099  
QY 106 CCGCTGCGCGCGCATCCGAAACCGCGCCAGGATCAGTCGACGACCGCGATGCGCG 165  
DB 1100 TCCCGCGCGCGCGCGCGCGCTCCCTGCTGCTGCTGCGCGCGCGCGCGCGCGCG 1159

QY 166 CTTGATCTCGAGCGTGGCGGCGAGCTCGATCTGTCGTCGCGCGCTCCACACCGAGGGAA 225  
 Db 1160 TGGCCCGCTCGGACCGAGCGCTGCGCCGCCACCCCGCGCGCGCGCAAGTCAGTCGT 1219  
 QY 226 ACTGGCTCGGTGGCAGCGGCGAGCCGCCAGCGGACAGCTCGTCCAGGCTGGCGTCAAGGA 285  
 Db 1220 GCGGGTGGCAGCGGCGAGCGGACAGCGGCGGCGCGCGGCGGAGCGCGCGCGCG 1279  
 QY 286 CTTTCGCGGGTGGGAGCGGAAACGCGACGCGTCAATCCCTGTCAGTCCGATCGGCT 345  
 Db 1280 GCGCAGCGCGGTGTCGTGGCTGGGAGC-CCCTGAGCCCCGGCGCGGTGGCGCGCC 1338  
 QY 346 CAGTGGCGGCTGCTCCCTTGGCTGGGAGGATAGCGGTTCACGACGAGCGGACACCGG 405  
 Db 1339 CTGAGCGCTCGCACACCTGAGCGCGGTGCGGCGCGCGCGGAGCGGAGCGTGGCGCTG 1398  
 QY 406 CGGCGGGGGCGGGGGGTTTACCGATCGCTCGATGACCGAGCGGCTGCGGGGTCGGG 465  
 Db 1399 CTGCGCGAGGGGCGCGCTGCAACACAGCGCGTCCATGTCCATGCCCGTGGCGCACTCG 1458  
 QY 466 CGGTGCGGAGATCCGTACCGCGCGGACCGCTCGGCGCGCGCGCGCGCGG 515  
 Db 1459 CCGCGCGCGCACCGCGCGCGCGTCCCTGCTGCTCCAGAGCGGCGCACGG 1508

## RESULT 25

AL162497/c

LOCUS

DEFINITION Human DNA sequence from clone Rp11-313L9 on chromosome 13, complete sequence.

ACCESSION

AL162497

VERSION

AL162497.20

KEYWORDS

HTG.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

Mashreghi-Mohammadi, M.

Direct Submission

TITLE

Submitted (05-JUN-2001)

JOURNAL

CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT

On Jun 8, 2001 this sequence version replaced gi:14280409.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations.  
 Together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em, EMBL; Sw,  
 SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP  
 database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)  
 This sequence was generated from part of bacterial clone contigs of human  
 chromosome 13, constructed by the Sanger Centre Chromosome 13  
 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr13>  
 Rp11-313L9 is from the library RPI1-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBAC3.6  
 IMPORTANT: This sequence is not the entire insert of clone  
 Rp11-313L9 It may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.

The true right end of clone Rp11-313L9 is at 143409 in this  
 sequence. The true left end of clone Rp11-40B6 is at 100074 in this  
 sequence. The true right end of clone Rp11-358F13 is at 100 in this  
 sequence.

## FEATURES

source

Location/Qualifiers

1. 143409  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /clone="Rp11-313L9"  
 /clone\_lib="RPI1-11.2"  
 1. 350  
 /note="L1M9A repeat: matches 5421. .5789 of consensus"  
 351. .654  
 /note="L1PB1 repeat: matches 5846. .6155 of consensus"  
 655. .957  
 /note="L1M9A repeat: matches 5789. .6085 of consensus"  
 2063. .2106  
 /note="23 copies 2 mer tg 91% conserved"  
 2063. .2106  
 /note="11 copies 4 mer ttgtg 93% conserved"  
 2138. .2422  
 /note="L1MD3 repeat: matches 6691. .7023 of consensus"  
 2422. 3051  
 /note="L1MC4 repeat: matches 7168. .7837 of consensus"  
 3674. .4545  
 /note="L1PA4 repeat: matches 5272. .6144 of consensus"  
 5701. .5910  
 /note="7 copies 30 mer 89% conserved"  
 6490. .6613  
 /note="MIR repeat: matches 92. .225 of consensus"  
 9599. 9648  
 /note="25 copies 2 mer aa 94% conserved"  
 10251. .10559  
 /note="AluX repeat: matches 2. .312 of consensus"  
 11134. .11521  
 /note="L1MB7 repeat: matches 5762. .6161 of consensus"  
 12667. .12710  
 /note="11 copies 4 mer ttgtg 81% conserved"  
 13210. .13309  
 /note="L2 repeat: matches 2559. .2674 of consensus"  
 13794. .14148  
 /note="THE1C repeat: matches 1. .371 of consensus"  
 14816. .15030  
 /note="MIR repeat: matches 49. .256 of consensus"  
 14982. .15037  
 /note="L2 repeat: matches 2651. .2705 of consensus"  
 15476. .15940  
 /note="L2 repeat: matches 1314. .1835 of consensus"  
 16305. .16609  
 /note="AluX repeat: matches 1. .300 of consensus"  
 16683. .16722  
 /note="20 copies 2 mer tt 77% conserved"  
 16948. .17073  
 /note="MLT1C repeat: matches 348. .464 of consensus"  
 17074. .17360  
 /note="AluJ repeat: matches 13. .299 of consensus"  
 17361. .17661  
 /note="MLT1C repeat: matches 7. .348 of consensus"  
 17974. .18347  
 /note="L1MD1 repeat: matches 5659. .6029 of consensus"  
 18399. .18554  
 /note="3 copies 52 mer 75% conserved"  
 18400. .18549  
 /note="5 copies 30 mer 73% conserved"  
 18413. .18552  
 /note="5 copies 28 mer 74% conserved"  
 18565. .18738  
 /note="L1MD2 repeat: matches 6150. .6335 of consensus"  
 19070. .19105  
 /note="9 copies 4 mer gata 88% conserved"



repeat\_region 19363..19664  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 20032..20053  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 21272..21554  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 22754..22918  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 26601..26684  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 26685..26814  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 26815..27110  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 27111..27285  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 27286..27578  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 29945..30257  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 30812..31078  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 31897..32091  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 32818..32853  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 32856..32957  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 32958..33406  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 33407..33867  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 33895..34718  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 33915..34519  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 33920..34103  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 33924..34553  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 33925..34730  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 33940..34709  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 33954..34213  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 33999..34698  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 34014..34650  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 34071..34642  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 34300..34709  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 34390..34689  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 34488..34699  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 34512..34703  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 34629..34718  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 35062..35647  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 35694..35885  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 35882..36229  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 36377..36484  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 39838..39877  
/note="L2 repeat: matches 2436..2750 of consensus"  
repeat\_region 39841..39875  
/note="L2 repeat: matches 2436..2750 of consensus"

repeat\_region /note="7 copies 5 mer tggtt 85% conserved"  
41466..41514  
repeat\_region /note="L2 repeat: matches 2699..2747 of consensus"  
41792..42143  
repeat\_region /note="THE13 repeat: matches 1..364 of consensus"  
43471..43590  
repeat\_region /note="L2 repeat: matches 2629..2750 of consensus"  
44937..45157  
repeat\_region /note="L2 repeat: matches 2259..2489 of consensus"  
45157..45157  
  
Query Match 10.5%; Score 54.8; DB 9; Length 143409;  
Best Local Similarity 46.2%; Pred. No. 17;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
  
QY 46 ACTGCCACTCCGGCGGGTGACACAGTCCAGTGGAGTTGAGCAGCCAGCTCATCG 105  
DB 124992 AGTTCCGGCGCGCAGTAGACCCATCGTCGGGTCTGCGCCACGACCCCATCAGCG 124933  
QY 106 CCGCTTCGCCCGGCGATGCCGAAACACCGGGCCAGGATCACGTGACGACGCGGATGCGCG 165  
DB 124932 TCCCGCGCGCGCGCGCCACACACCTGCTCACTGCCCCCAGCAGCGGCGCTGG 124873  
QY 166 CTTGATCTCGACGGTTCGGCGCGAGTCTGTCGCCCGGCTCCACACACAGGGAA 225  
DB 124872 TCGCGCGCTTCGGCGCGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 124813  
QY 226 ACTGGCTCGGTGCGCAGCGCGACGCCCGCGGACAGTCTGTCAGGTGGCGTCAACGA 285  
DB 124812 GCGCGGTGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 124753  
QY 286 CTTCTCGCGGTTCGGGAGCGGAAACGCGCAGCGCTCAGATCCCTGTCAGTGCATCGGT 345  
DB 124752 GCGCAGGCGCGGTGTCGGTGGAGC-CCCTGAGCCCCCGCGCGCGCGCGCGCG 124694  
QY 346 CAGTCCCGCTCGTCCCGCTTGGCCTGGGAGGATAGCGTTTCAACGACGCGCGCACCGG 405  
DB 124693 CTGAGCGCTCGCACACCTGAGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCG 124634  
QY 406 CCGCGCGCGCGCGCGCGCGGTTCAGCGGATCCCGTTCAGTACGAGCGGTGCGCGGTGCGG 465  
DB 124633 CTGCGCGCGCGCGCGCGCGTGCACACAGCGCTCCATGTCCATGTCGCCGCGCACTCG 124574  
QY 466 GCGGTTCGGGAGATCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 515  
DB 124573 CCGCGCGCGCGCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 124524

RESULT 26  
AC138109/c 224777 bp DNA linear HTG 16-JAN-2003  
LOCUS Mus musculus chromosome UNK clone RP24-444115, WORKING DRAFT  
DEFINITION SEQUENCE, 8 unordered pieces.  
AC138109  
VERSION AC138109.2 GI:27764789  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 224777)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (13-DEC-2002) Genome Sequencing Center, 4444 Forest Park  
REFERENCE 2 (bases 1 to 224777)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (16-JAN-2003) Genome Sequencing Center, 4444 Forest Park  
REFERENCE 3 (bases 1 to 224777)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (16-JAN-2003) Genome Sequencing Center, 4444 Forest Park

**COMMENT**

STREPTOMYCES COELICOLOR A3 (2)  
SOURCE  
ORGANISM

AL093312.1 GI:27443764  
VERSION  
KEYWORDS

[illegible]

ORIGIN

Query

```

COMMENT      On or before Oct 26, 2002 this sequence version replaced
              gi:20520847, gi:20520774, gi:20520824, gi:20520691, gi:20520851,
              gi:20520775, gi:20520830, gi:20520831, gi:20520832, gi:20520833,
              gi:20520835, gi:20520841.
FEATURES             Location/Qualifiers
     source          1..30800
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     RBS             84..89
                     /note="973"
     gene            98..973
                     /gene="SCO2256"
                     /note="synonym: panB"
                     /note="973"
     CDS             98..973
                     /gene="SCO2256"
                     /note="SCC75A.02, panB, 3-methyl-2-oxobutanate
                     hydroxymethyltransferase (EC 2.1.2.11), len: 291 aa.
                     Highly similar to many including: Escherichia coli
                     SW:PANB_ECOLI(EMBL:L17086) 3-methyl-2-oxobutanate
                     hydroxymethyltransferase (EC 2.1.2.11) (264 aa), fasta
                     scores opt: 711 z-score: 839.4 E(): 0.44.3% identity in
                     262 aa overlap and Mycobacterium tuberculosis
                     SW:PANB_MYCTU(EMBL:Z70692) 3-methyl-2-oxobutanate
                     hydroxymethyltransferase (EC 2.1.2.11) (281 aa), fasta
                     scores opt: 1019 z-score: 1199.2 E(): 0.59.1% identity in
                     269 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="3-methyl-2-oxobutanate hydroxymethyltransferase
                     (EC 2.1.2.11)."
                     /protein_id="CAB61702.1"
                     /db_xref="GI:6469241"
                     /db_xref="GOA:Q9RKS2"
                     /db_xref="SWISS-PROT:Q9RKS2"
                     /translation="MTQLSRAQTPQPKFADGNRALYGGKINRITVDITAAKRGEXK
                     WPLMTAYDAMTASVDGSEIGVLMVGDSPAGNCHIGYETTVTIDEMTAAVVRGT
                     SRALIVGLDFPGSQEGEVQALRSATRLVKAGAVLKEGERSHQIIELVESGIP
                     VMAHIGLTPQSVNSGRVQGRGEAAQQLLRDAKAVQDAGAFVLELVPALAAEV
                     TRTLHIPTVGIAGGPDITDAVLVMTDMLGLTGGRMPKFKVQYADLRKVMGDAAKAYAE
                     DVVGSTFPADSHSVH"
     gene            1093..2190
                     /gene="SCO2257"
                     /note="synonym: SCC75A.03"
                     /note="973..2190"
                     /gene="SCO2257"
                     /note="SCC75A.03, probable ABC transporter, ATP-binding
                     component, len: 365 aa. Highly similar to many ABC
                     transport proteins e.g. Streptomyces peucetius
                     SW:DRRB_STRPE(EMBL:M73758) daunorubicin resistance
                     ATP-binding protein (330 aa), fasta scores opt: 922
                     z-score: 1009.0 E(): 0.48.9% identity in 329 aa overlap and
                     Streptomyces antibioticus TR:Q53716(EMBL:L06249)
                     ATP-binding protein OLEC-ORF4 involved in
                     oleandomycin-resistance (325 aa), fasta scores opt: 1883
                     z-score: 2452.8 E(): 0.89.7% identity in 329 aa overlap.
                     Also similar to several Streptomyces coelicolor putative
                     ABC transporter ATP-binding components e.g
                     TR:CAB42771(EMBL:AL049841) SCE9.25C (315 aa), fasta scores
                     opt: 1031 z-score: 1127.7 E(): 0.54.4% identity in 318 aa
                     overlap. Contains Prosite hits to PS00017 ATP/GTP-binding
                     site motif A (P-loop) and PS00211 ABC transporters family
                     signature. Also contains a Pfam match to entry PF00005
                     ABC_tran, ABC transporter."
                     /codon_start=1
                     /transl_table=11
                     /product="probable ABC transporter, ATP-binding
                     component."
                     /protein_id="CAB61703.1"
                     /db_xref="GI:6469242"
                     /db_xref="GOA:Q9RKS1"
                     /db_xref="SPTREMBL:Q9RKS1"
                     /translation="MAPAGCRWDVGALSVPPAIVGMKRIIDIPRAADSAVTVRGLVK

```

```

HYGETKALDGVLDVRECTVMGVLPNGAGKTLVRLITLITPDSGOAHVAGYDVVR
QPRQRRVIGITQTASVDEKLPQWENLYMIGRLLDLSRRDARTRADLLRPSLTER
KGRPASTYSGEGRRLDLAASMI GHPAVLFDDEPTGLDPRTRNEVWDEVKRMWDGV
TVLTYQMEBAEGLASLTVDKRGKVIANGIEELKAKGVGRTRLIRPADPLQRLPL
AAYLDLGITGLASSTVDTERTGTVLPILSDEQLTAVVGAVTARGITVSAVTTLPSSL
DEVLSLTGTHRASAPQDPAPVTDTRREAVV"
1276..1821
/gene="SCO2257"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 189.70, E-value 4.5e-53"
1297..1320
/gene="SCO2257"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
1594..1638
/gene="SCO2257"
/note="PS00211 ABC transporters family signature"
2174..2180
/gene="SCO2257"
2187..3011
/gene="SCO2258"
/note="synonym: SCC75A.04"
2187..3011
/gene="SCO2258"
/note="SCC75A.04, probable ABC-transporter, transmembrane
component, len: 274 aa. Similar to many transport proteins
including: Streptomyces peucetius
SW:DRRB_STRPE(EMBL:M73758) daunorubicin resistance
transmembrane protein (283 aa), fasta scores opt: 523
z-score: 593.2 E(): 1.2e-25 35.1% identity in 245 aa
overlap and Streptomyces antibioticus
TR:Q53717(EMBL:L06249) membrane protein OLEC-ORF5 involved
in oleandomycin-resistance (273 aa), fasta scores opt:
1502 z-score: 1685.2 E(): 0.83.2% identity in 274 aa
overlap. Contains multiple possible membrane spanning
hydrophobic domains."
/codon_start=1
/transl_table=11
/product="probable ABC-transporter, transmembrane
component."
/protein_id="CAB61704.1"
/db_xref="GI:6469243"
/db_xref="GOA:Q9RKS0"
/db_xref="SPTREMBL:Q9RKS0"
/translation="MSATYAPDTLTTPDARITPRAHLRHTSALIRNLLWIROPES
MFDAILFPVITLLFVYFVGGSIGQSGGQQAQYQVYIPGLMAMGMNAQGVGTGF
NODFNSGVNDFRSLPIGRGSVLFKIAVELLMFLFATVLMVAVLVGFDITDWAGL
LASVGLSAGVSGALMWVLTGLGVNMKQSVQMGFLVLMPLQFGSSIFAPDTSMPGW
LOSFTDYNPLSLADTARGMLGCGPVADHLVTLWSAALTAVMAPIAIHKFRTKS"
complement(3033..7037)
/gene="SCO2259"
/note="synonym: SCC75A.05c"
complement(3033..7037)
/gene="SCO2259"
/note="SCC75A.05c, possible multi-domain regulatory
protein, len: 1334 aa. Similar in parts to many
regulators. Similar in the N-terminus to Streptomyces
coelicolor SW:AFSR_STRCO(EMBL:D90155) regulatory protein
AfsR (993 aa), fasta scores opt: 524 z-score: 356.0 E():
1.9e-12 37.0% identity in 316 aa overlap. Similar in the
C-terminus to many other proposed regulators e.g.
Mycobacterium tuberculosis TR:O53720 (EMBL:AL021931)
transcriptional regulatory protein (1085 aa), fasta scores
opt: 809 z-score: 541.8 E(): 8.3e-23 34.1% identity in 883
aa overlap. Contains a Prosite hit to PS00017
ATP/GTP-binding site motif A (P-loop). Also contains a
highly degenerate repeat region with 16x PA(P/S)G and 5x
APG. This repeat region also has a notably higher GC
content (83.4%). Contains probable coiled-coils between
residues 867 to 904 (38 residues) and 1093 to 1121 (29
residues)."
/codon_start=1
/transl_table=11
/product="putative multi-domain regulatory protein."

```

/protein\_id="CAB61705.1"  
/db\_xref="GI:6469244"  
/db\_xref="GOA:Q9RKR9"  
/transl="SPTREMBL:Q9RKR9"  
LALRAGVAVPAGLLVEEVNAGPPADATGALQALVGRRLRAGLADSVASAGGVYELAA  
APDDVHLRFDLDELADGDPAAVAVLDALALNRPVLPDLPTDRTAARW  
ETRHEALRARTALDGOAHSLELPELTLDSPLDCHPLDEPLQALRLRDRGRTAE  
LAAVEAVRLADRLLGTPDPLRLTHAELLSPLSPPTTGRSTPGMTSGPAGAG  
AAGTGVASGAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG  
TASADPAGPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGT  
PAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGT  
GRPTVAGPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGTAPAGT  
VERLRADLAATLVLGGGAGKRLSCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG  
EAVLAVAGARETVLRAGBAGAAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG  
DAARLAELALRCPLTALREPLVGPESLREPLPELPEALRLRLLDGAARAP  
GFRVDADETAACAEICRLDGLPLAELAAELRLMPLPQIADRLDRFLRTSGS  
PLPQOQLRAVWMSMDLDADEVLRLSVFAGGCDLAAAEVAGCPAALDSLGS  
LVDSILVAVPTDRLSGDGMRYLLETVAEVAGERLDEAGGAAARAHLYTYRELA  
RTDPLRGRPLALIERREVENLRTALHAVAERDEQELCALCLALSLVWYWMRD  
RVERNFVWVAGLPPDPABPTGTPARPWERTSAPPMTCEVLAERGVHLALA  
CMDTELDWONPAQRKRVIAADTYEPMPQTCSSPGLNFTYVNLGMDRLRRIMD  
ATVTRCTRETPGYEWELAGLQMRANLANRTDWDADVADDESLEIYDRUGDAGWA

Query Match		10.5%; Score 54.6; DB 1; Length 300800;
Best Local Similarity		47.7%; Pred. No. 15;
Matches 234; Conservative		0; Mismatches 244; Indels 13; Gaps 2;
QY	26	GCCCCGCGCCCATCAGCAATCGCCATCGCGCGGGGTGACAGTCCACAGGTGGA 85
Db	6399	GCCGAGTTCGCGAGGCGCGCAGCGCGCGCGCTCGAGCGGGTGGCC 6458
QY	86	GTTGAGCAGCAGTCTATCGCCGCTCGCGCGCATGCCGAAACACCGCGCGCAGATCAC 145
Db	6459	CTGCAAGCGCGGTGAGTCTGGGAGGAGTCTCGGCTGGCGAGGTCCAGGCGGC 6518
QY	146	GTCAGCAGCGGATCGCGCTCGATCTCGAGCGTGGCGCGGAGTCTCGATCTCGTGGCC 205
Db	6519	GGTGTGCGCGCGCGCGAGGCTTCGAAGTGGCGGCTCTCCAGCGCGCGCGCTCGCGCGT 6578
QY	206	CGCTCCACACACGAGGGAACGTGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 265
Db	6579	GCGTGGCGAGGTGCGCGA-----GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6626
QY	266	GTCAGCGTGGCTGCGGACCTCTCGCGGTGCGGAGCGGAAACCGCGCGCGCTCAGAT 325
Db	6627	GTCGAGCAGCAGCGCGCGCTTGGCGGGTGGCGGCTCGCGCGA-GGGCGCGGGTCCCT 6685
QY	326	CCCTGTGAGTGGCTGAGTCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 385
Db	6686	CACGCGGAGCGGTCGAAAGCGTGGAGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 6745
QY	386	CAGACAGCGCGCACACCG 445
Db	6746	AACGCGCGTGGCGAGGCGACGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 6805
QY	446	CCAGCGGCTGGGCGTGGGCGCGTGGCGAGATCCGTACCGCGCGCGCGCGCGCGCGCG 505
Db	6806	GCGTTCGAGGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6865
QY	506	GCGCGCGCGCG 516
Db	6866	GCAGCGCGCG 6876

RESULT 28  
AF386507/c  
LOCUS  
DEFINITION Streptomyces lavendulae complestatin biosynthetic gene cluster,  
complete sequence.  
ACCESSION AF386507  
VERSION AF386507.1 GI:15029316  
KEYWORDS

SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

Streptomyces lavendulae  
Streptomyces lavendulae  
Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomycetes.  
1 (bases 1 to 55972)  
Chiu,H.T., Hubbard,B.K., Shah,A.N., Eide,J., Fredenburg,R.A.,  
Walsh,C.T. and Khosla,C.  
Molecular cloning and sequence analysis of the complestatin  
biosynthetic gene cluster  
Proc. Natl. Acad. Sci. U.S.A. 98 (15), 8548-8553 (2001)  
21353013  
PUBMED 11447274  
2 (bases 1 to 55972)  
Chiu,H.T. and Khosla,C.  
Direct Submission  
Submitted (28-May-2001) Department of Biological Science and  
Technology, National Chiao Tung University, 75 Po-Ai Street,  
Hsinchu 30050, Taiwan, R.O.C.  
Location/Qualifiers  
1. 55972  
/organism="Streptomyces lavendulae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1914"  
2020. 3086  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator"  
/protein\_id="AAK81822.1"  
/db\_xref="GI:15029317"  
/translation="MMVRGSTDTRIVDVRALSAGSPNGGPNPDHVAALAEIDL  
PLPPIVHRMSRVIDGLHRLRAAQRHSTIAVTFYDGTEDAFVLAVESVNRHGLP  
LSUPDKRAAVRIATHPQNSDKIASVTGIATVADVVRSPAPIEPADFPARGID  
GRAPVDIASRGRAGQLISRNPLSLRIARAAAGSPETARDVNRRLSREGPELPP  
RPGVRPREELKVVPQSGEERHDVAGFTEGLSAVNRKLADPALRLNETOR  
QLRLSLVHMLPPAEWEGIVDTPVYMSGTVAHLARECAHIEWEVAARVEKVAQGD  
SPSRADSSACPTV"  
3491. 5617  
/note="ORF2"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter"  
/protein\_id="AAK81823.1"  
/db\_xref="GI:15029318"  
/translation="MLVTQPGVMKGGDDSPSKIRPGTSRRVLAFLVPHRKLATA  
LIFVLVDSLVIVTTELLLEIVDSGIMKDDASLVVLLVALLVLAALAGSVTOLAQA  
LSGRIGGVSDLVQRVQAFHVQRMPLAFFRTQTGTGALSRLSDTVVMAQALITLLS  
ATGATVVLVLAEMFVLSVLSIALVFFLFFMFLYIGRRLOAQRKQKQKELHNL  
GTHERTNVQAGLAKLFGPAELADYQGEAEVREVRLTIFSSLAFLMALLTA  
LTTALVYGGGLVDGVLGTLVATLGRIFGPTITQSSQANALVLSFDR1  
FELDKPLTEERPDVALPAPSASAPETTRFDQVSPAYPRPEVSLASTHSAAG  
ORSPKPVHLGLDPSVAGTTLALVGPSCAGKTTIHLVSRLYDVTSGAVRIDKQV  
RDLTFSLRATGVVAQADALFHDITRANLAYARPEATEBELIACEQARLIDRSI  
PDGDTVVGDRGFRFGKQRLAVARLLKAPAIVVLDDEATHLDESEANAQORALK  
TALSNRTSLVIAHLSTVREADQLIVDDGRVREKQTHDELIAGGLYTHLNTQFSK  
SETLPPRSNGHREQAAPQGTGPRITRGTGAGGGMGGPGGLGGGMPGGGMPGGMG  
GGGMPGGGMPGGGPGTGRVVRGDSGSP"  
5664. 12014  
/note="ORF3: comA"  
/codon\_start=1  
/transl\_table=11  
/product="peptide synthetase"  
/protein\_id="AAK81824.1"  
/db\_xref="GI:15029319"  
/translation="MNGTSVAAPVSLLEFRARVVEAPEAVVVEGGRVSYSGELDAA  
SDRAVAALRDREVGRDRAVRLERSADLIGVLGVWKGAGAAVVPDGVPAERVAVF  
LKNSTPAVTDVAVGAGDASSVALDGDLLAYVMYTSGTPTKGVAVPHGSVAALVG  
ESGGLGPDSDLVFHAPEADISLFEVWPLASGARVIAEPGVAVDAAVRRHIAAG  
VTHVHTAGLFRVLAEEAPDCTGAREVLGGDVVPLEAVRVRACCPDVVRHLYGP  
TESTICATWHLVEFGDDTARVLPIGHPLTNHIIHVDLSDLRVAPGTGVELYTAGGL  
ARGYLKAGLSAERFVACFPADGERMTDGLARWTDGDELAFAGRADQVKIKGFRV  
ELGEVEAALAAQPAVAVVAVREDRPEKELVGLVDPDQADSDVLRARLAQTLPD  
YMWPAAFVWLIDALPLTVNGKVDHKAPEPTATTGREPTPAETLCELFAVLGGE  
HVGADSFELGGDSIMSQLAARGNRKIAPSAQDVFEHETPAIAAQLVETPDS

CDS  
CDS  
CDS

AAPOVDAPLADLTADAEAELEALVPEGLAEIWPITLPLQEGLLFHAAYDEQGPVYEGQM  
TLESGLDVARLAAWNGVAARAGTIRAGFHELRLASGLVQGVVADVAIPMRVEDLSG  
LGAEITLAEATAAQOTERFDPARPILLRLVLVMSDERHLIIVVTHLIDGNSM  
SVLDEVAALVAGSDGALSAPSPEDHLAWARODKEAALAWEALDGVDEWTLV  
DPQERAKSDAGLTPREPHQFTARLIAAFARAGGLTYNTLLOGANALVLRITGSTD  
VIGAPVAARTLPEFGVGVGVFNMTVPVRVFDGQVRVDMITLALQKQSGALISGH  
HVGUSELQKUTGSSASVGVFNMTVPVRVFDGQVRVDMITLALQKQSGALISGH  
LADHLQVRLAHRLAESEADAGELHALTVLEQVVDVAFMPVGLSVFVGAADAP  
LPDTPAARTSLPFPAPFTQARTATATLAKAARLTYLAELASHDLHLAHLAP  
GIGDTPVALPAPSVBELITALLIGALAGGAYVFPDPHADRIAVFLADAAVAVLC  
TAQTRPMVPEFTGLVTLDELPEDGPDGGRVLTAPADHIAVITLSTGCTPKGA  
ATWGLRLNLVADRIEYGIQTRVLQVLPSPDSMDIWPITLCAEGLVLAIPGG  
HTTGDEGLDLADHRIITHANVPAVOLTHLPDRBLPALRVLSGDLAPADTRRRWVAR  
CDLNEGVTEATVVSITAPLDAGPLTTGGPIARAGVHLVDGPRVPPVGTGELY  
VTGQVAGYLNRLTHFAHRIADPRTFGGRMTYTGDLVRHTHGGELVFGRADEQVK  
LRGHRIGELVEAALADHPDQVAAAHIDARLITGVVADGRVPDFSALEASRML  
PDMPLAAVVLDELPTNPKLDREALPODPASEAGRAPRTPAEVFPCTAVAEPTL  
GLRVGADDSFVLLGDSITSLQVTRARQAGLVVTPROIPDEKTPERIALVAEPTL  
SGRESADDDGVGEVMPAAHFTGERVAGGAFACQMTIATPPGLRQDALATALSALVE  
THAMLRARVDDAGPLIAGGDSVUTAGILARVADATGTPGALADWTEAAREAVR  
LDPGAGAMVADDDGPRVGRVLVLAHHLAVDGVSWRLIVLPDKETYEAAAGHSP  
RPAEVPVTSYXRRTLLAAEAHSDHRIAEPLAMPALLRDATLFPVPPDPARDTAATV  
RHTGWTLRNDTATVGRTAALFHCGIHDLIATLAGAVHMRDDYAPVNVVEGHG  
RTPDGLADLSRTGWTFSAPHVRLDVSGVDLFGAMDGERAAGLLIKSVQIIRAPVD  
GLGGLRLYLNAETAPVLAELPLPOGVFNGLRFPDARAGGARVGPWEPATPRALYS  
SDDPRMPLPHALEAGAVVTRHCPETLTLSRPAVLADDAADSLGRINWALLRGLA  
AHSRRPAVGCTPSPDFLLDLQDEVEQLEAIAADLEGGLSR"

CDS

/note="ORF4; comb"  
/codon\_start=1  
/transl\_table=1  
/product="peptide synthetase"  
/protein\_id="AAK81825\_1"  
/db\_xref="GI:15029320"  
/translation="MTSTTPVHGSALEAVNPLSPLOEGLLPHAEYGGQGPDLTIORT  
DLLEGPDLADLRASWALLNRHLSALRSPRRKSGRTVQLI PREVALNRHRELDSDT  
DLTEEAELADLRASWALLNRHLSALRSPRRKSGRTVQLI PREVALNRHRELDSDT  
VPMVLTAAIYAADGASGLGAPTPFRDLHAWLRQDRSAGEAMRKEALADGPTL  
MAPADPHSPVPRHGLTLDSELDALTTLARRHGLTVNTVMQAWAMLLARLTGRT  
DVLGATVAGPASPGLPVESMVLFINSVVRVRLDADQPLLLTLEQARQSALLPH  
QHLSPLEIQRVGGGAFADAMLENYPLPREPAAPCALAFHLEGRQATHYALITG  
SVPGQLHYTPADLITDVAQDLIGRLWLEQLVSDPSVLVGRVGLVGRGLRGL  
VVEGNATAGVPVSGSVLENFRARVAQAEVAVDVGQVSYGDELDAENRMAAPYL  
QGWGGRVAVVLESLIDILIAALLGVKAGAAVVPVDSAYPAERVAFVMDVSAPYL  
TIDDPVSVTAEGEVEVETAGDIAYVMTSGTGTGKGVAVPASHVAALVPPGFGV  
GPDVAPLPHAPADVLSLFEVWPLASGRIVVAEPMVAVDGAARHFIADGVTHVH  
TAGFLVLAESAFCDFGVRHPLNGVDVPLEAVERVRAACPDVVRHLYGPTVSLC  
ATMHLFPEGEQGLVPLGRPLNNRVYVLDPLQPPVPGVTGELVYVAGLAGLARGVIG  
RAGLSAERFVAPSPADSERMYRTGDLVRWTTGVELVFGVGRADQVLRGFRVELGEV  
AALAAQPAVAAQVAVVAREDPGEKRLVGLVPSGEEPTEAVHASLADRLPEVMPAA  
LVVLDALPLTVNGVYDHKALPAEFTATASREPTAAKLLCELFPAELVGLERVGAD  
SFFELGDSIMSMOLAAAGTRRGVYVFAQDVFEYETPAATAAVALAEADGEAARDVGT  
GPVPTPALAASPAAAPGPAQWVVGAPPVLDVDTLITAGLVGLDTHMLRARVYD  
GPRLVVGRDGVDPSSLVSVLRVTDPCGELDGAVERAARAAGRLDPAAGVWVRLVVD  
AGPLVGLVLAHHLVVDVSWELLPLDIAEACEAAQGEAGEPGHAGKGEQAD  
CAGFEFVGTGFRWALLESAREERVELDQWRTVLAGDPAICGRALNPARTPA  
GLTERHAWTPASQARVIVGRTPGVFGHGVHEVILLATLAAVNEWRTGTDHDLVVDV  
GHGRTAVSAGELSRVGTFTSVPRVIRDLADIGADVADAGAAAGALLKAVKQSRV  
NDPGLVGLLHLPDTPASVLAQPLTPQGVFNLYGRFPAPATEVAAPQAPGATLGG  
NDPMPVYKHALDAGAVIRDSAGPDLTLLSHSPGRLDTASVQDLGRINWALLRGLA  
HTEPDSAGGHTASDFLLRLDQDEVELEAEFADEHP"

CDS

16705\_31401  
/note="ORF5; comC"  
/codon\_start=1  
/transl\_table=1  
/product="peptide synthetase"  
/protein\_id="AAK81826\_1"  
/db\_xref="GI:15029321"  
/translation="WTRSQVEDVMPLSPLOEGLLPHAEFSDGPDVYOGORPLDLGCP  
VDADLRTSWALLTRHVLVLAGFRPKRSGEAVQVIAREVELPWRMEDLSLEDEGTAA  
AKLAELAEARQNRNLNAPPLRLLLVRLAENRHLIIVTSHHIIIMDGMWNPVLLHGR  
TAVYAGADGAAPKPPSYRDLVLAWLQNRDAREARWRELLGAEPVLLADPGR  
TPVLPESLIADIPEKLTIGELTGLARQHGLTVNTLITQAWAMVLAARLGRDTRVVEGATV

AGRPAGLAGVSMIGMFINTLPVRVELDGERPSVEMLMKALQTRSALIAHQHGLAEI  
ORLGAGAVFTLVVFENYPRPPEAPADFTIRFAAQSTAHYPLTLVAVPKDRLM  
FKLDVDPDLFRETAEAVFGRILVLEQLAARDPLQVGRVADGAEARHTIDWNST  
ATDRPAESVLELFAQVVRTPHATISDGERQLTYAOLAVSGLARHLADRGVSAGD  
RVAVULERSADLLITLAVKAGAAVFPDTPAERIAFLLTCDPAAVVCSSRTV  
VVPEDVPAPLPDLEASLIDGEPISLPVAGDIIAYMYTISSTGVKGVAVPHGSVAA  
LVGEPGAVPGDSVLPHAFADISLFEVWVPLASGARVEAPSVVDAVAAREYI  
AGGTVHVTAGLFRVLAEPSEFECTGAREVLTDGDDVVPARAVRACPEVVRHL  
YGTPEVSLICATKHLLEPEESPQVLVGRPLNGRQVYLDAYLHVPASVGTGELYIAG  
AGLAVGLKRAGLSERFVACFPADGERMYRTGLVRMTSDGELVFGRAADAQKVRG  
FRELGEVEAALAEAGVHAUTAREDPGERLIGVLPDGEDVTLDEHLEHAKV  
LPEVMVPAALVLDALPLTANGKVDHAKALPAPDPAVRVTGREPTDAERTLEICEIAAV  
LGLERVGTDEFFHLGDSITSQWVAVARAGRADLVLTPROVEDLATPRTAARAPAA  
TRITTPEDAGVCELPWTPMARGRAARPGFQAVVAAATGALGELTTLTALGTLID  
THMLRLAAADATPTLVVARGSVDAALVARVELTDEDPEVVEAARAAGADLDP  
AAGVMRLVWDAGPVRAGRLVTVHHLVVDGVSWRILLPDLRACEAARAAGLADL  
PVGTPEPAHQLEAEAREDEKRIAEALDRWELLGDGTLIGARALDPAARDTAATLRH  
SWTLSGAPAHDLIGRISAAFCGVHEALLATLARAVARSRGDASAVLDVVEGCHREP  
LASVDLSRTVGMFPADAPHLVDADIDVADASAGAAAGALLKAVKQSRVAPGDGL  
GYGLLRHLNPDTSALAQIPAPQVGFNYLGRFPAPPTTAAAGSPAGSGSIADPTLLH  
TVAAAVVVDTPKGPETLITLNWAGNILDADAERLGGTLTALLITGLATHLITDPSAGG  
HTPSDFLADLRHSWNTLVARHDMRVSPHELDGRTVQAVTGAEPWREVDLSDR  
LDLTGFLDADLRHSWNTLVARHDMRVSPHELDGRTVQAVTGAEPWREVDLSDR  
PEPELAABAEALAEAREAFDLPTRAPLLRLFLVLRGRRHLVMTSHIILDDGWSPT  
IALDEASRIYAADGASGLGAPSPRDLHAWLRQDKDTAHEAMRAELAGADEPTLIA  
PTACRGAPAAPEGDSLLSSELTALTSVAREHGLTNTLAQAGWALVLSRLARLTDV  
VGIYVAGSPPELPGVESVMFNTLPVRVRLDGAESAVRMLARLQDRQSALLAHQH  
VGLAETISLGGAGVFTMLMFPYFGGPHLAATDDVTVERVNTRAGTSYPLAVGI  
MEGDMRHHVYTRPDLFRALHVARQVVERVLEQVADPAAPVGRVVDVGPGLRGLV  
VPGWNTTTRTLSSVLEFRFAVQAQADPAVAVDGERQVSYAELDAESNLAAHLRE  
RGVGRDRAVRLERSIDILIAALLGVKAGAAVVPDGAIPAERVAFVLEDSAPVIT  
DAVVGGEAEVSAVTEGDIAYVMTSGTGTGKGVAVPHGSVAAVLEGTGMLGPGD  
SVLFHAPHADISLFEVWVPLASGARVVAIEPAGVADAARVHRTAAGVTHVHTAGL  
FRELADPECFNGSAHEVLTGDDVPLEAVARVRAACEPVARVHLVGPTEVSLCATMH  
LLEBEGDTRVLVPGSPLANQVYVLDLAFLOPPVPGVAGELYIAGAGLARGVLRAGL  
SABRVACFPADGERMYRTGDLIARDTDDGELAFGRADAQVAKIRFRVLEGEVEAALA  
AQPAVAAVAREGPRKEKRLVGLVPHGQSPDSVLRARLAQTLPYMVAALVLA  
DALPTVNGKVDHKALEPFTATASREPTAAEKLVCLEFADLVNLRQVADSGFFE  
LGGDSIMSMOLAAAGTRRGVYVFAQDVFEYETPAATAAVALAEADGEAARDGGE  
VPMVLTAAIYAADGASGLGAPTPFRDLHAWLRQDRSAGEAMRKEALADGPTL  
MAPADPHSPVPRHGLTLDSELDALTTLARRHGLTVNTVMQAWAMLLARLTGRT  
DVLGATVAGPASPGLPVESMVLFINSVVRVRLDADQPLLLTLEQARQSALLPH  
QHLSPLEIQRVGGGAFADAMLENYPLPREPAAPCALAFHLEGRQATHYALITG  
SVPGQLHYTPADLITDVAQDLIGRLWLEQLVSDPSVLVGRVGLVGRGLRGL  
VVEGNATAGVPVSGSVLENFRARVAQAEVAVDVGQVSYGDELDAENRMAAPYL  
QGWGGRVAVVLESLIDILIAALLGVKAGAAVVPVDSAYPAERVAFVMDVSAPYL  
TIDDPVSVTAEGEVEVETAGDIAYVMTSGTGTGKGVAVPASHVAALVPPGFGV  
GPDVAPLPHAPADVLSLFEVWPLASGRIVVAEPMVAVDGAARHFIADGVTHVH  
TAGFLVLAESAFCDFGVRHPLNGVDVPLEAVERVRAACPDVVRHLYGPTVSLC  
ATMHLFPEGEQGLVPLGRPLNNRVYVLDPLQPPVPGVTGELVYVAGLAGLARGVIG  
RAGLSAERFVAPSPADSERMYRTGDLVRWTTGVELVFGVGRADQVLRGFRVELGEV  
AALAAQPAVAAQVAVVAREDPGEKRLVGLVPSGEEPTEAVHASLADRLPEVMPAA  
LVVLDALPLTVNGVYDHKALPAEFTATASREPTAAKLLCELFPAELVGLERVGAD  
SFFELGDSIMSMOLAAAGTRRGVYVFAQDVFEYETPAATAAVALAEADGEAARDVGT  
GPVPTPALAASPAAAPGPAQWVVGAPPVLDVDTLITAGLVGLDTHMLRARVYD  
GPRLVVGRDGVDPSSLVSVLRVTDPCGELDGAVERAARAAGRLDPAAGVWVRLVVD  
AGPLVGLVLAHHLVVDVSWELLPLDIAEACEAAQGEAGEPGHAGKGEQAD  
CAGFEFVGTGFRWALLESAREERVELDQWRTVLAGDPAICGRALNPARTPA  
GLTERHAWTPASQARVIVGRTPGVFGHGVHEVILLATLAAVNEWRTGTDHDLVVDV  
GHGRTAVSAGELSRVGTFTSVPRVIRDLADIGADVADAGAAAGALLKAVKQSRV  
NDPGLVGLLHLPDTPASVLAQPLTPQGVFNLYGRFPAPATEVAAPQAPGATLGG  
NDPMPVYKHALDAGAVIRDSAGPDLTLLSHSPGRLDTASVQDLGRINWALLRGLA  
HTEPDSAGGHTASDFLLRLDQDEVELEAEFADEHP"

Query Match		10.4%;	Score	54.2;	DB	1;	Length	55972;
Best Local Similarity		52.4%;	Pred.	No.	26;			
Matches		119;	Conservative	0;	Mismatches	108;	Indels	0;
Gaps		0;						
Qy	97	AGCTCATCGCGCTCGCGCGCATGCCGAAACACCGGGCCAGGATCACGTGCAGCACCG	156					
Db	33554	AGCACTCCGGTCTCTCTCGGCAGCACCGGAAAGGTGCCCGCGTCACTGCAGGGCGG	33495					
Qy	157	CGATCGCGCGCTCGATCTCGACGGTGGCGCGCAGCTCGATCTCTGTCGCCCGGCTCCCA	216					
Db	33494	TCACCCCGCGCGCGACCTCTCTCGCGATCCGCGAGCGCGTCCGACCGCCGCGCTCGCCA	33435					
Qy	217	CGAGGGGAAACTGGCTCGGTGGCAGCGCGAGCCCGACCGCGAGCTCGTCCAGGCTG	276					
Db	33434	CGACCAAGCGGGGCGCGGAGACCGCGCGCACCGACATCTCGAACACGAGGGCTCGAAG	33375					
Qy	277	CGTCACGACCTCTCGCGGCTCGGAGCGGAAACCGGCAACCGCTCAG	323					
Db	33374	CGTGGGGCGTGCAGCAGCACCGCGTCTCTGAGCGTGCACGCGCCAG	33328					

RESULT 29  
AF429315/c  
LOCUS  
125020 bp  
DNA  
linear  
PRI 18-JAN-2002

```

DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
AUTHORS Holmes S.E., O'Hearn E., Rosenblatt A., Callahan C., Hwang H.S.,
Ingersoll-Ashworth R.G., Fleisher A., Stevanin G., Brice A.,
Potter N.T., Ross C.A. and Margolis R.L.
TITLE A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes S.E., Ingersoll-Ashworth R.G., Ross C.A. and Margolis R.L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES
source
1. 125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chrnosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/notes="isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
complement(336507..>36887)
/genes="JPH3"
/notes="synonym: JP3"
complement(336507..>36887)
/genes="JPH3"
/product="junctophilin 3"
complement(336507..>36887)
/genes="JPH3"
/notes="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translation="MSGGRFNFDDGSGVCGHEDKAGHGVCTGPKQGGYTSWS
HGFEVLGVYTPSGNTYQGTWAGQKRGIGLESKWKVYKGEWTHGFKRGYVRECA
NGARYETGWSNGIQDGYGTETYSDG"

repeat_region
10.4%; Score 54; DB 9; Length 125020;
Query Match 10.9%; Pred. No. 23;
Best Local Similarity 220; Mismatches 143; Indels 4; Gaps 2;
Matches 45; Conservative 220;

QY 44 GAATGTCACCTCCGGCGGGTGACAGGTGCAACAGGTGGAGTTGAGCAGCCAGTCAAT 103
Db 51141 SRWYSSMAWYKGGRRKTKSMATSCRMGMAMRSKGMRYKWSCRYKGMFWGRWSGY 51082

QY 104 CGCCGCTGCGCGGCGATCGGAACACCGGGCCAGGATCACTGAGCAGCAGCGGATGCG 163
Db 51081 KGCSGYMSMGYSGRSTKYCSYTGGYCSKSAAMCKSKSTSKCTKSKYSCS--MGRY 51024

QY 164 CGCTCGATCTCGACGGTCCGCGCGAGTCTGATCTCGTCCCGGGTCCACACACGAGGG 223
Db 51023 SSYCTCGWRCKMGKSSCCASRSRGSKYKMGASWRCRRGRGMSCTWSSMKSYSNAYK 50964

QY 224 AAATGCTCGGTGGCAGCGCAGCCCGGAGCAGCTCTGTCAGGTGGCGTCACG 283
Db 50963 SSARGRGMCKYKTSCKMYKYYSMHBKSHDBSGYSKYSRYBYWTCMMKDSNCWSVDSHM 50904

```

---

```

QY 284 GACCTCTCGCGGTTCGGAGCGGAAACGGCAGCGCTCAGATCCCTGTTCAGTGGCATCG 343
Db 50903 RDMYDKRMKVSSTGTRGVHBSAWMKSMRSDSGYSGMYHMKSYMKSSASKMYYSG 50844

QY 344 CTAGT--GCCGTCTGCCCTTGCCTGGAGGATAGCGTTTCAGCAGCAGCGCACC 401
Db 50843 MKSVVBMDCMSYTBESKSVRMRWSGYBYRKCAVRRYBHRSTSTSVYMGKRSC 50784

QY 402 ACAGCGCGCGCGCGCGCGGTTCAGCGATCCCGTCCGATCAGCAGCGCGC 453
Db 50783 WVRDSVSRRSVHTSMNRCRCRGARMAKSCAGMKMVCMTGCVHSRCS 50732

RESULT 30
PM11H12G/c 1393 bp DNA linear STS 29-MAY-2003
LOCUS Penicillium marneffe STS, clone pm11h12-g, sequence tagged site.
DEFINITION
ACCESSION AL684264
VERSION AL684264.1 GI:19338009
KEYWORDS STS
SOURCE Penicillium marneffe
ORGANISM Penicillium marneffe
REFERENCE 1
AUTHORS Yuen K.Y., Pascal G., Wong S.S., Glaser P., Woo P.C., Kunst F.,
Cai J.J., Cheung E.Y., Medigue C. and Danchin A.
TITLE Exploring the Penicillium marneffe genome
JOURNAL Arch. Microbiol. 179 (5), 339-353 (2003)
MEDLINE 22595073
PUBMED 12640520
REFERENCE 2 (bases 1 to 1393)
AUTHORS Danchin A. and Pascal G.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES
source
1. 1393
/organism="Penicillium marneffe"
/mol_type="genomic DNA"
/db_xref="taxon:37727"
/clone="pm11h12.g"

ORIGIN
Query Match 10.3%; Score 53.8; DB 11; Length 1393;
Best Local Similarity 43.7%; Pred. No. 76;
Matches 223; Conservative 0; Mismatches 285; Indels 2; Gaps 1;

QY 8 CACGAACGACCGGTGGTCCCGCGCGCCATCAGCACTCCGCGCGGGGTGAC 67
Db 1273 CCCGTGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1214

QY 68 CAGGTGACAGGTGGAGTTGAGCAGCAGCTCATCGCCCTCGCGCGCGCATGCCGAA 127
Db 1213 CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1154

QY 128 ACACCGCGCC--AGATCAGTCAGCAGCAGCGGATGCGGCTCGATCTCAGCGTCCG 185
Db 1153 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1094

QY 186 CGCAGCTCGATCTCGTCCCGCGGTTCACACACGAGGGAACACTGCTCGTGGCAGCGCG 245
Db 1093 CNNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1034

QY 246 AGCCCGACCGGGAAGTCTCGTTCAGGTGCGGTACGAGACTCTCGCGCGGTCCGGAGCG 305
Db 1033 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 974

QY 306 GAAACGCGCAGCGCTCAGATCCCTGTGAGTCGCTCAGTGCCTCGCTCGTCCCTT 365
Db 973 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 914

QY 366 GGCCTGGGAGGATACGGGTTTACGACGAGCGGACACCGCGCGCGCGCGCGCGGT 425

```





Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H.,  
Isenbarger, T.A., Peck, R.F., Pohlschrod, M., Spudich, J.L.,  
Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P.,  
Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and  
Dasarma, S.

Direct Submission  
Submitted (14-JUL-2000) Institute for Systems Biology, 4225  
Roosevelt Way NE, Seattle, WA 98105, USA

## FEATURES

## source

1. .14720

/organism="Halobacterium sp. NRC-1"

/mol\_type="genomic DNA"

/strain="NRC-1"

/db\_xref="taxon:64091"

complement (152. .1093)

/gene="VNG1350C"

complement (152. .1093)

/gene="VNG1350C"

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng1350C"

/protein\_id="AAG19683.1"

/db\_xref="GI:10580863"

/translation="MSFRPTLGETPTTVLFLLLAAIWSSFVAIKIGIADVPALISFA

AIKYSFAGILVLAAYTAHTDRWRPRGRREWLAAVGLFVIGIYHGALVYLGHDVSAP

IAKIVTSLAPITKCFASKLHENVITDQTAAGLAGLVGVIVANPGSAGSVDLGGI

LIVFLAASAPILGGVLTGKLPGGGLPVETMEAWMLIGSLALWGAVRGENTAVH

WPTALASLAYITLVCAGGGLFVILLERAGATENILVSYLEPIGASVVAWALFGDV

VSPRALAGFAVIFAGPALVKHRYIGAVENTTDDAEI

1236. .1721

/gene="trh6"

/note="synonym: VNG1351G"

1236. .1721

/gene="trh6"

/note="trh6"

/codon\_start=1

/transl\_table=11

/product="transcription regulator"

/protein\_id="AAG19684.1"

/db\_xref="GI:10580864"

/translation="MDRITITLKAIDLTGSPALOEETGIPVSTIHYRLNLRDD

GVINLDINPDSIGLDTVIVLVTSDGGYEDVADEISAVEGVQTFTFMGTDF

VVARLPGSDDDVERLISDEALDPVDRTNSTFVDRHKAQALSESYSLSLVTALAD

E"

complement (1729. .3597)

/gene="gatB1"

/note="synonym: VNG1352G"

complement (1729. .3597)

/gene="gatB1"

/note="gatB1"

/codon\_start=1

/transl\_table=11

/product="Glu-tRNA amidotransferase"

/protein\_id="AAG19685.1"

/db\_xref="GI:10580865"

/translation="WTEFYDELGLVAGLEIHQOLDTATKLCACPTTTEREPABADRT

FTYLPFTSELGIDEALERSVERFEYLAIDTTCVLEDDDPFHRDLDDALAA

LETHGLHMDVRAHVRKVIDSGNQFQSTWAEAGADTSEGVGIEDLMLA

ESQAQRIEDREYLSLDRLGIPVETGTFDISPAQAREARERIGMLRSTQKV

RGLGTIRQDVNVIADGARVEKMGVQSLDIDDLVREVRQVELLDIVDELDDAA

VGPRTVDVTFSGVIRGDLDDGGEVHAPLHGFDGLVGRKLQADRLRGTFSDDH

ATRHAGGIIFHTDELPAVGTVAAEVAALRDVAGGEDDAVAIVADDPETAQSIQAVA

ERATAMAGVPESTRANDGTGSKYLRPLPGARWYPETDVPDPDSAVETPELLT

EKVRYQADFDLGLAGQVAYGRWOLFQQVQVAGVADATLAAQLESITVLSRRDDV

FVRLDTHFRVGLVADGLAQEGVPBELLAALAPQPSDPAVLAELGLGSAADE

VREAVGVVERNSDQVAEGMGAFSALMGCEMGALRGADGDLNSELVREBEIQQRs"

complement (3634. .5178)

/gene="VNG1353C"

complement (3634. .5178)

/gene="VNG1353C"

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=11  
/product="Vng1353C"  
/protein\_id="AAG19686.1"  
/db\_xref="GI:10580866"  
/translation="MKLHPLSVPRALSGFMAFLIGGAALSGNAALTCGAVLGG  
VMVAVVAYEYAYVRRDYELTSDLIASGVFSREREIPLRVQNDVTRSLVA  
RLGLAVVETAGGSGSTASLNFVSREAAADLQADVTRTRTPGDAADAATAATDA  
GGDSRSRAADRAEPTPVGERGELLFALSNRDLGLYGLSFPNPLFGIITAVTA  
APSLGRMAVPDYGVMVLVAGVALGVMVLSAAVVFQFYGRFLRRVGGDDURVRG  
LFRDGTTPLSKQTVSYEDSVLMERYGFAVLAVETAGYAPQSPSGSSEAAVPIAP  
RAAVFLAREIEPVADIDLSRPPERAQDRYRYALAGLVVAGVVAAGLVSRLVVSFPWY  
ALAVLPLAIPAARAHATRGVYSDSDHVVTQAGWWRKTTVPAHRVOTLVKQTVF  
QREWALTSVIDTAGSRSLGGGAVALDRDGGDADALADDVVDCTLLSVGVRGDDTAA  
AGSAAD"  
complement (5175. .5654)  
/gene="VNG1355H"  
complement (5175. .5654)  
/gene="VNG1355H"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="vng1355h"  
/protein\_id="AAG19687.1"  
/db\_xref="GI:10580867"  
/translation="MERLDARVVVWMLRVFVAVVLAGAAGTGGALDVGVSAAAA  
VAVGVVALLGATHALLRYRWRFDVREDTLYIEHGLVVRVTRTPVYRVQHVDSRRG  
PLERVLGGRVVVYTAGSGADVTIPOLSRERAADAQETLALAISEPESGGDAV"  
complement (5699. .7111)  
/gene="fumC"  
/note="synonym: VNG1356G"  
complement (5699. .7111)  
/gene="fumC"  
/note="FumC"  
/codon\_start=1  
/transl\_table=11  
/product="fumarate hydratase"  
/protein\_id="AAG19688.1"  
/db\_xref="GI:10580868"  
/translation="MSEDYRTEQDSLGEQVVPADAYWGAQTORAIENFPISGIAFQRR  
FVRALGVKKAQAQRDLGDDERADAIVAAADEVIAGEDDDQFPDVFQSGTGS  
SNNAENVIANRAAELLGEGIDRVVHPNDVNVGSSNDVIPTAMHVASLDALVNDV  
KPGLETLAELDKADAFQGVKTGRTHLQDAPVRLGQFGYGVYQVKEGIDRIEAV  
APRLSELALGTVAGTGLNTHPEFTAGYISEETGVTFREADNHEFAQAADHANE  
AHGRLTGVGNLKNKIANDLKLLASGFRANGLGEIQEQNPQSSIMFGKINPVVAEAVN  
QVHKVQVGNDAIAAGAAEQGIDINLYKPVLAHNFLOSADMLANASAAFGKEKFAKLE  
ANEAACEAQRVSRMALATALNPTIGYDKASEVAKAALKEGKTIVTEVVVEKYLSEABA  
ADVLDPGEMTHRGILSGDDT"  
complement (7315. .7563)  
/gene="VNG1357C"  
complement (7315. .7563)  
/gene="VNG1357C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1357c"  
/protein\_id="AAG19689.1"  
/db\_xref="GI:10580869"  
/translation="MLADIEARIEAGFEADCSATYSRARGPHDDDDHLEAVVSPAFE  
GVSLVDQHDVYDALGDAMTEITHALELITRPSDRDA"  
complement (7612. .8595)  
/gene="rad2"  
/note="synonym: VNG1359G"  
complement (7612. .8595)  
/gene="rad2"  
/note="rad2"  
/codon\_start=1  
/transl\_table=11  
/product="DNA repair protein"  
/protein\_id="AAG19690.1"  
/db\_xref="GI:10580870"  
/translation="MGNADLRQAAIEETPFADLEGSVAVDAHNMILYKYLTTTVOVT  
GADVYTTSDGTEVANLVGAQGLPKPKEFTHGLITPVFVWDGGVTEKLDKADREQR  
YEQLDDAREAGAAEAARDARTQRTPTTHETTRFLLDLPQVEAPAEQAQA



YMRTRDADVAGSDYDCLLGSPTVTLRLQTLSSGHPMLDFDIALAEHDLTWEQLVD  
VGLTCGTDFNPGIDTGFPTTALDAIGEHGDLMDVLAAGEHVAHGDRIRLEFLNPDT  
DEVVDPDVSPTDADARAFTVDEWEVDADAVARGFERIDAAAQTLGDRWT"  
complement (8644..9138)  
/gene="VNG1360H"  
CDS complement (8644..9138)  
/gene="VNG1360H"  
/notes="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="vng1360h"  
/protein\_id="AAG19691.1"  
/db\_xref="GI:10380871"  
translation:"MNTGKAVAYEGAPLAKRGDDWPPDARAADDVDDVGVAVFNE  
DRDADAAWLYVTGDRGAGIGARLCAFAAAHLLGDHCVRIAVNPPFAYEALHK  
AGFGTGOETGIAELVLRCEHRAAGYQDLDRYERDLSPEEAFLDRGASPPA  
RVSW"  
9315..9716  
/gene="VNG1362H"  
CDS 9315..9716  
/gene="VNG1362H"  
/notes="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="vng1362h"  
/protein\_id="AAG19692.1"  
Query Match 10.2%; Score 53.2; DB 1; Length 14720;  
Best Local Similarity 48.7%; Pred. No. 52;  
Matches 145; Conservative 0; Mismatches 153; Indels 0; Gaps 0;  
Qy 2 GTGCAGCAGCAACACCGGTGTCGCCCGCGCCCATCACGAACCTGCCACTCCGCGCG 61  
Db 3934 GTACACGCCCGGTGTCGCCCGCGCCCATCACGAACCTGCCACTCCGCGCG 3993  
Qy 62 GGTGACCAAGTCGACCGGTGGAGTTGAGCAGCAGCTCATCGCCGCTCGCGCGCAT 121  
Db 3994 GGCCAGCGGTACACAGGGAACGACACGAGCGGCTCACGAGCGCGCGCGCCAC 4053  
Qy 122 CCCGAACACCGGCGCAGGATCAGTGCAGCAGCGCGCTCGATCTCGAGGT 181  
Db 4054 CACGCCACGCGCGGACCGGTACCGCGGACGTAGCGGTCTGAGCGCGCTCGGCGG 4113  
Qy 182 CGGCGGAGCTCGATCTCGTCCCGGCTCCACACAGGGAACCTGGTCTGGTGGCAG 241  
Db 4114 CGGCGAGGTGCGATGTCGCGACCGGTTCATCTCCGCGCGAGCTCGAACCGCGC 4173  
Qy 242 CGGCGACCCCGCGGAGACCTGCTCGAGCTGGCGTACGACCTCTCGCGGTG 299  
Db 4174 CGGCGCGGATCGGACCGCGCTCGACCGCGCTGCGACTGCCCCGCGCG 4231  
RESULT 33  
AP004323 95209 bp DNA linear HTG 21-MAR-2002  
LOCUS Oryza sativa (japonica cultivar-group) chromosome 6 clone  
DEFINITION OJ118 C02, \*\*\* SEQUENCING IN PROGRESS \*\*\*.  
ACCESSION AP004323  
VERSION AP004323.1 GI:16930108  
KEYWORDS HTG; HTGS PHASE2  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
1  
REFERENCE  
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.  
TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC  
clone:OJ118 C02  
JOURNAL Published Only in Database (2001)  
REFERENCE 2 (bases 1 to 95209)  
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.  
TITLE Direct Submission

JOURNAL Submitted (14-NOV-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai  
2-1-2, Teukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/.  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
COMMENT The nucleotide sequence of this BAC clone was generated by  
combining Monsanto and RGP-Japan sequencing data.  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
Location/Qualifiers  
1..95209  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="6"  
/clone="OJ118\_C02"  
ORIGIN  
Query Match 10.2%; Score 53.2; DB 2; Length 95209;  
Best Local Similarity 45.4%; Pred. No. 32;  
Matches 178; Conservative 0; Mismatches 212; Indels 2; Gaps 1;  
Qy 107 CGCCTGCGCGCGATCCGAAACACCGCGGAGGATCACGTGACAGCGGATCGCGCG 166  
Db 69303 CGACCGCGCGACCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 69362  
Qy 167 CTCGATCTCGAGGTGCGCGCGAGTCTGATCTGTCGCGCGCTCCACACAGGGGAAA 226  
Db 69363 GGGGGCG 69422  
Qy 227 CTGCGTCTGTCAGCG 286  
Db 69423 GCGGCG 69482  
Qy 287 CTCGCGCGGTGCGGAGCGGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 346  
Db 69483 CCG 69542  
Qy 347 AGTCCCGGTCTGCG 406  
Db 69543 CCG 69602  
Qy 407 GGGCGCGCGCGCGCGCGGTTCAGCGCGATCCGCTCGATGACACGCGCGCTCGCGGGT 466  
Db 69603 CCG 69660  
Qy 467 CGGTGCGCGAGATCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 498  
Db 69661 CCG 69692  
RESULT 34  
AE016920 305584 bp DNA linear BCT 14-OCT-2003  
LOCUS Chromobacterium violaceum ATCC 12472 section 11 of 16 of the  
DEFINITION complete genome.  
ACCESSION AE016920  
VERSION AE016920.1 GI:34104104  
KEYWORDS Chromobacterium violaceum ATCC 12472  
SOURCE Chromobacterium violaceum ATCC 12472  
ORGANISM Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Chromobacterium.  
REFERENCE 1 (bases 1 to 305584)

**AUTHORS**  
**CONSRMT**  
**TITLE**  
**JOURNAL**  
**MEDLINE**  
**PubMed**  
**REFERENCE**  
**AUTHORS**

Brazilian National Genome Project Consortium  
The complete genome sequence of Chromobacterium violaceum reveals  
remarkable and exploitable bacterial adaptability  
Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)  
2382980  
1450082  
2 (bases 1 to 305584)  
Vasconcelos, A.T.R., de Almeida, D.F., Almeida, F.C., de  
Almeida, L.G.P., de Almeida, R., Gonçalves, J.A.A., Andrade, E.M.,  
Antônio, R.V., Araújo, J., de Araújo, M.F.F., Filho, S.A., Azevedo, V.,  
Batista, A.J., Batista, L.A.M., Batista, J.S., Belo, A., vander Berg, C.,  
Blaney, J., Bogo, M., Bonato, S., Bordignon, J., Brito, C.A.,  
Brocchi, M., Buriy, H.A., Camargo, A.A., Cardoso, D.D.P.,  
Carnielli, N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C.M.,  
Falcão, B.S., Chueire, L.M.O., Faria, I.B.C., Duran, N., Fagundes, N.,  
Falcão, C.L., Fantinatti, F., Farias, I.P., Felipe, M.S.S.,  
Ferrari, L.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R.,  
Freitas, N.S.A., Furlan, L.R., Gazzinelli, R.T., Gomes, E.A.,  
Gonçalves, P.R., Grangeiro, T.B., Grattapaglia, D., Griseard, E.C.,  
Guimarães, C.T., Hanna, E.S., Hungria, M., Jardim, S.N., Laurino, J.,  
Leão, L.C.T., Passarella, L., Lima, A., Loureiro, M.F., Lyra, M.C.P.,  
Macedo, M., Madeira, H.M.F., Manfio, G.B., Maranhão, A.Q.,  
Martins, W.S., de Mauro, S.M.Z., de Medeiros, S.R.B., Meisner, R.V.,  
Mench, C.F.M., Moreira, M.A.M., Nascimento, F.F., Nicolas, M.F.,  
Oliveira, J.G., Oliveira, S.C., Paixão, R.F.C., Parente, J.A.,  
Pedrosa, F.O., Pena, S.J.D., Perreira, J.O., Perreira, M.,  
Pinto, L.S.R.C., Pinto, L.S., Porto, J.I.R., Petrich, D.P.,  
Neto, C.E.R., Reis, A.M.M., Rigo, L.U., Rondinelli, E., dos  
Santos, E.B.P., Santos, F., Schneider, M.P.C., Seunear, H.N.,  
Silva, A.M.R., da Silva, A.L.C., Silva, D.M., Silva, R., Simões, I.C.,  
Simon, D., Soares, C.M.A., Soares, R.B.A., Souza, E.M., Souza, K.R.L.,  
Souza, R.C., Steffens, M.B.R., Steindel, M., Teixeira, S.R.,  
Urmenyi, T., Vettore, A., Wassem, R., Zaha, A. and Simpson, A.J.G.  
Direct Submission

**TITLE**  
**JOURNAL**  
Submitted (22-JAN-2003) Labinfo, LNCC - Laboratório Nacional de  
Computação Científica, Rua Getúlio Vargas 333, Petropolis, RJ  
25651070, Brazil

**FEATURES**  
**source**

1. 305584  
Location/Qualifiers

/organism="Chromobacterium violaceum ATCC 12472"  
/mol\_type="genomic DNA"  
/strain="ATCC 12472"  
/db\_xref="taxon:243365"  
370. .2154  
/locus\_tag="CV2796"  
370. .2154

**gene**

**CDS**

/locus\_tag="CV2796"  
/notes="identified by sequence similarity; putative; ORF  
located using Blastx/COG0006"  
/codon\_start=1  
/transl\_table=1  
/product="probable peptidase, M24 family protein"  
/protein\_id="AAQ60464.1"  
/db\_xref="GI:34104105"

/translation="MTIPQRTASRAAKNAGADACLVPSDDPHISYLPQWQARQ  
WLSGFTSGMTLIVGADFAGLWADSRVYQAEQELAGTQIMKIDTAASSLHQLWA  
DNRCTQTLVDGDLGLAAKALQALPEPGVIVRADLDDAVVAGRPALPAAYV  
QHDGDPASRAKLARVREMGADRHFIITLDDIWLNLGAGVSYNPFIAH  
ALITDNATLFWAPGKIDALAAQLAADGVADYAAKPAALSPALGALLIDPRVT  
LGRVAVDRVRIEINSTEIMKRSAAAEHVRAMEQDGAALPEAFVAFANVN  
RARIITLIDDEITAAARPPGVSFGTIAFNANGALPHYATPEAHSEIQDGL  
LITDGGQVGLGTTDITRVVAVGTPSAQKRDFTLVILKGTMAISMHPHPRGTSPLD  
ALARAFLWHDIDFGHTGTVGVFLNVHEGQPSIRAVPEPHMAQEGMITSIEPGV  
YRAGQGVRIENLVANVSRNQFGFLKFEITLCPIDRRCIDLSLLTQPEIAMNG  
YHAEVRARLPLVDGARDWLLANTETL"  
2225. .3118  
/locus\_tag="CV2797"  
2225. .3118

**gene**

**CDS**

/locus\_tag="CV2797"  
/notes="identified by sequence similarity; putative; ORF  
located using Glimmer/GeneMark/Blastx/COG0583"  
/codon\_start=1

/transl\_table=1  
/product="probable transcriptional regulator, LysR family"  
/protein\_id="AAQ60465.1"  
/db\_xref="GI:34104106"  
/translation="MPFRLADRLRFQDASQGSFSAARXKHLSPAAASACIQMWEA  
ALNARLFQTRRLRAGETFLAYCRQALEEGRHQQAQQLSGVRLSAPS  
DLGRNRLDCLDRFGAHPAVHSLSLDTADLIDDDIDLAIRYQGSADSLVARQL  
AASRVCAAPSLLEKLGQSHPODLAGLPTLTLTGHPNMWRYRUKGELRSRLLE  
RVQNSDNCVLEWRWAVQGHGFKSRLEIDADLAAGRLTTLVDDYFIEPAPLHLLYPG  
HRLQPARIRLIDFLDQPA"  
complement (3474. .4790)  
/locus\_tag="CV2798"  
complement (3474. .4790)  
/locus\_tag="CV2798"  
/notes="identified by sequence similarity; putative; ORF  
located using Blastx/COG0477/TC:2.A.1.21.1"  
/codon\_start=1  
/transl\_table=1  
/product="probable macrolide efflux protein"  
/protein\_id="AAQ60466.1"  
/db\_xref="GI:34104107"  
/translation="MGAAFAFCMLSEMITGLSLLQLFALGVWYIQKTSVEDFSFAL  
LNGTLPAFLPIAGACADLLDRRKVLIIACDFLLLLIGVLSFMIWLDALLVPAAMVQ  
NGVALIHLFRPSVQVALSRLLPVNQLSOANGRLTSLDGLVQLAGPLMAGLMFQFG  
LTSILPLETLLAAVLLVIRAMRGNTPEAAAGNTVGRMLRHSRSLSKTMYLNGQR  
AMFGLLIYMLLQALITLVSSMTPLVILARHGSDDTIGVMGCGAFKAIGATVITSN  
PERLMRWIISDLLAVAMAGFVEKPLWALAAAMACGSVNSCTLSLWMTK  
PLDWRGSLFSLIASINLAASAMLLAGVILVERVLAPAMPGGLADLLGWLIGTQGG  
RGVALLVLCGSLAISVALLGNTLAQLDDMVEDAPICLSESA"  
complement (4817. .6136)  
/locus\_tag="CV2799"  
complement (4817. .6136)  
/locus\_tag="CV2799"  
/notes="identified by sequence similarity; putative; ORF  
located using Glimmer/GeneMark/Blastx/COG3208"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAQ60467.1"  
/db\_xref="GI:34104108"  
/translation="MTVKRMVQDSVSGAALASLEGEGTGRMLWLAQPEHGQTQLOG  
LAYALAEVRLLQDSFYRLAGVGLAYEVLQGRMGQTVACVLEGAPVWVRQ  
SEALAVERSQALSPAWRLCLSYVPEPGDFLOPTNPASAEFWALGQAASARVP  
AHQPLVMRGROSLPPLICPGAGASVVDFFALSALAYQGVSYQPRGLDGRPP  
HVTVEDAARSYLDALANAPGVHLLGHSGFWIAHLEAAQLQAGTRVLSLTVLDSR  
VPNDAGEVQOREVLSELIALYQGGEGEPLRLDAELASREASQKALHGLLRYGI  
LPSRSSPDILSGVIRTFACALRAGYRPERLYQGSRLVLVVKPSEASSEATSEQWI  
AGRRRLPDLVFWEGFNHITLRSFQAGHLAAWLDMLFDMKAVHS"  
complement (6199. .6627)  
/locus\_tag="CV2800"  
complement (6199. .6627)  
/locus\_tag="CV2800"  
/notes="identified by sequence similarity; putative; ORF  
located using GeneMark/Blastx"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAQ60468.1"  
/db\_xref="GI:34104109"  
/translation="MALRMGRPAGRLFGVWALGILLASACVGNKLEIDVDEPYVKQ  
ALSREGTKQWDLFQCPQRKISFDGNEIWIYSYRQYKAKLQNLTNVSVLRGQNN  
YSKELVLFQSKVRKWLKASEESVGTGFAIQRPAEN"  
complement (6644. .7489)  
/locus\_tag="CV2801"  
complement (6644. .7489)  
/locus\_tag="CV2801"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAQ60469.1"  
/db\_xref="GI:34104110"  
/translation="MRSAGFGLAVWLTSALAAGTAAADLASSRTYAVLQKGAVFQFE  
NAXLAPNGVYQYRLYLNGVSNASAPTLKLSRGGSKAVPLGARQVFLSPITAEKSDAS



```
misc_feature 82587..84446
/note="ORF 29; positive strandedness"
misc_feature 84481..85548
/note="ORF 30; positive strandedness"
misc_feature 85556..86845
/note="ORF 31; positive strandedness"
misc_feature complement(86803...87372)
/note="ORF 32; positive strandedness"
87494..88420
/note="ORF 33; positive strandedness; N-terminus only"

ORIGIN
Query Match 10.2%; Score 52.8; DB 6; Length 88421;
Best local Similarity 46.5%; Pred. No. 38;
Matches 240; Conservative 0; Mismatches 272; Indels 4; Gaps 2;

QY 7 GCACGACGACCGGTGTGCGCCCGCGCCATCAGAACTGCCACTCCGGCGGGTGA 66
Db 67518 GCACGACGACGTCGCGCCCGCGCGGATGAGCGCGCTGCGCGCGAGCGGA 67577
QY 67 CCAGGTGCACAGGTGGAGTTGAGCAGCAGCTATCGCCGCGCTGCGCGCGATCCGA 126
Db 67578 ACTGCCCGCGGTGCGCGCGAGCCGCTGCGCGCGGTGCTGCTGCTACCGCGA 67637
QY 127 AACACCGCGCGCAGGATCACGTGCAGCAGCGGATGCGCGCTCGATCTCGACGTCGGCC 186
Db 67638 CGGCTCGCGCGACCTGTGTGTGGCCCGCGCGACCGCTGACCGCGACGCTGAT 67697
QY 187 GCAGCTCGATTCCTGCCCGGTCTCCACACGAGGGAATCGCTGCGTGGCGAGCGGA 246
Db 67698 CGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67757
QY 247 GCGCCAGCGCGGACAGCTGTCAGAGTGGCGTCAAGGACCTCTCGCGGTGCGGAGCGG 306
Db 67758 GCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67817
QY 307 AACCGCGCAGCGCTCAGATCCCTGTAGTCAGTCGCTCAGTCCCGGTGCTCCCGCTTG 366
Db 67818 CGAGCTCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67875
QY 367 GCGTGGGAGGATACGCTTTCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 424
Db 67876 GCGAGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67935
QY 425 TTCAAGCGATCCGCTCGATGACACGCGCGCTGCGGGGTGCGGGCGGTGCGCGAGATCCGTA 484
Db 67936 TTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67995
QY 485 CCGCGCGGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 520
Db 67996 GCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 68031

RESULT 36
AC069466/c 192634 bp DNA linear HTG 06-SEP-2000
LOCUS Mus musculus clone RP23-394J16, WORKING DRAFT SEQUENCE, 23
DEFINITION Mus musculus clone RP23-394J16, WORKING DRAFT SEQUENCE, 23
unordered pieces.
AC069466
AC069466.2 GI:9972319
VERSION HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 192634)
AUTHORS McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Dedhia,N.N., de la
Bastide,M., Huang,E.N., King,L., Kirchoff,K.A., Miller,B.,
Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Rodriguez,M.A.,
Shah,R.S., Shekher,M., Spiegel,L.A., Teth,X. and Vil,M.D.
Mouse Genomic Sequence
Unpublished
TITLE Mus musculus clone RP23-394J16, WORKING DRAFT SEQUENCE, 23
JOURNAL (bases 1 to 192634)
REFERENCE 2 (bases 1 to 192634)
```

AUTHORS  
TITLE  
JOURNAL

McCombie,W.R.  
Direct Submission  
Submitted (01-JUN-2000) Lita Annenberg Hazen Genome Sequencing  
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring  
Harbor, NY 11724, USA  
On Sep 6, 2000 this sequence version replaced gi:8134854.  
----- Genome Center  
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor  
Laboratory  
Center code: CSHL  
Web site: http://www.cshl.org/genseq  
Contact: mcombie@cshl.org  
----- Project Information  
Center project name: RP23-394J16  
Center clone name: RP23-394J16

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 23 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 30565: contig of 30565 bp in length  
\* 30566 30917: gap of unknown length  
\* 30918 52496: contig of 21579 bp in length  
\* 52497 52848: gap of unknown length  
\* 52849 71268: contig of 18420 bp in length  
\* 71269 71620: gap of unknown length  
\* 71621 83193: contig of 11573 bp in length  
\* 83194 83545: gap of unknown length  
\* 83546 93802: contig of 10257 bp in length  
\* 93803 94154: gap of unknown length  
\* 94155 103828: contig of 9674 bp in length  
\* 103829 104180: gap of unknown length  
\* 104181 113354: contig of 9174 bp in length  
\* 113355 113706: gap of unknown length  
\* 113707 121968: contig of 8262 bp in length  
\* 121969 122320: gap of unknown length  
\* 122321 129392: contig of 7072 bp in length  
\* 129393 129744: gap of unknown length  
\* 129745 136687: contig of 6943 bp in length  
\* 136688 137039: gap of unknown length  
\* 137040 142803: contig of 5764 bp in length  
\* 142804 143155: gap of unknown length  
\* 143156 147963: contig of 4808 bp in length  
\* 147964 148315: gap of unknown length  
\* 148316 153054: contig of 4739 bp in length  
\* 153055 153406: gap of unknown length  
\* 153407 157827: contig of 4421 bp in length  
\* 157828 158179: gap of unknown length  
\* 158180 162507: contig of 4328 bp in length  
\* 162508 162859: gap of unknown length  
\* 162860 167154: contig of 4295 bp in length  
\* 167155 167506: gap of unknown length  
\* 167507 171735: contig of 4229 bp in length  
\* 171736 172087: gap of unknown length  
\* 172088 175937: contig of 3850 bp in length  
\* 175938 176289: gap of unknown length  
\* 176290 180015: contig of 3726 bp in length  
\* 180016 180366: gap of unknown length  
\* 180367 184069: contig of 3703 bp in length  
\* 184070 184420: gap of unknown length  
\* 184421 187865: contig of 3445 bp in length  
\* 187866 188217: gap of unknown length  
\* 188217 191451: contig of 3235 bp in length  
\* 191452 191802: gap of unknown length  
\* 191803 192634: contig of 832 bp in length.

FEATURES  
Location/Qualifiers

1..192634  
/organism="Mus musculus"  
/mol\_type="genomic DNA"



gene	WPRG"		IDALLAGYRVFIIEVSPHPVLIPALRECFEAEVAAATVTLRRDQGGPPQVQARALGD	
	complement (6768. .7493)		GFVAGLAVDMSRWFVGDREAGDEGHRPRTVELTPTFFQRRYMLVAGVQVLPVAVLWVWALIA	
	/gene="nanr1"		VGTFRPAGALLSSAVELADGLVLSGRLLPGDAWVGAHTVAGVQVLPVAVLWVWALIA	
CDS	complement (6768. .7493)		ADRAGASLEILLRRLAPLEISGPSISSEAGVLAOVAVCAPDESGRELRILSRPAD	
	/gene="nanr1"		ADAGECHTCHANGSLAPCGPPADTGTATVPHPPAGAEALDPAGLYERERGVYGV	
	/note="putative regulatory protein"		AGAGGCTCHANGSLAPCGPPADTGTATVPHPPAGAEALDPAGLYERERGVYGV	
	/codon_start=1		PALRGVVALWRDGLVDVADLPEERAGGEGGADGDTAGTGLHPVLLDAALQFALL	
	/transl_table=11		AEPDGTGEGAGPEARLWLPFAWSGVRMLWATGARAARVRLSPLDGGGADVADERELRI	
	/product="nanr1"		AEPDGTGEGAGPEARLWLPFAWSGVRMLWATGARAARVRLSPLDGGGADVADERELRI	
	/db_xref="GI:31044141"		LSAVRGVAAALRAWVAEPRLLTVSRLLAVVTRGVAAGAGGEPVDDLAAGGACGARGVQ	
	/translation="MGLRVRVDEKSGSPRAQKLAVALVILIRSDQVMAQLTREI		AEHPDR LVLDVDDVDMGADTTDICAAGAAALGEPOVALRGDITLLAPRLARSAA	
	/note="putative regulatory protein"		TPGSAVDPNGTTLVTDSSGSLAGSVAEHLVRAEGRVHLLLVRFEGADGAYDTVDROD	
	/db_xref="GI:31044142"		AQMHTVDRDRTHALERVVAQVDPAPHLTGUVHVAGLSADIETSGHARGWAGVAVV	
	/translation="MGLRVRVDEKSGSPRAQKLAVALVILIRSDQVMAQLTREI		RFLGWAADLGGAGDGGRRVTGLGADRGALLRAACRADRPRLVAAIDIRTRALTAHDA	
	/note="putative regulatory protein"		HELPAALRTILGASASASAGGRAPVRRVAAAAPGRTTDMASRLVGLGPAERRAVLELV	
	/db_xref="GI:31044143"		SARAVLSEELGLEPSAALKELHNAVLVDAPLELEAA"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		8919. .17627	
	/note="putative regulatory protein"		/gene="nanr1"	
	/db_xref="GI:31044144"		/gene="nanr1"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/gene="nanr1"	
	/note="putative regulatory protein"		/note="PKS"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42855.1"		/protein_id="NAP42855.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42856.1"		/protein_id="NAP42856.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42857.1"		/protein_id="NAP42857.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42858.1"		/protein_id="NAP42858.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42859.1"		/protein_id="NAP42859.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42860.1"		/protein_id="NAP42860.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42861.1"		/protein_id="NAP42861.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42862.1"		/protein_id="NAP42862.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42863.1"		/protein_id="NAP42863.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42864.1"		/protein_id="NAP42864.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42865.1"		/protein_id="NAP42865.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42866.1"		/protein_id="NAP42866.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42867.1"		/protein_id="NAP42867.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42868.1"		/protein_id="NAP42868.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42869.1"		/protein_id="NAP42869.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42870.1"		/protein_id="NAP42870.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42871.1"		/protein_id="NAP42871.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42872.1"		/protein_id="NAP42872.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/db_xref="GI:31044144"		/db_xref="GI:31044144"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/db_xref="GI:31044145"		/db_xref="GI:31044145"	
	/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED		/translation="MRYEILGLRVRVDRSVTSISARKIEVLLAALLVRADEVVSED	
	/note="putative regulatory protein"		/note="putative regulatory protein"	
	/codon_start=1		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="nanr1"		/product="nanr1"	
	/protein_id="NAP42873.1"		/protein_id="NAP42873.1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	

AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.  
TITLE Medicago truncatula BAC Clone mth2-33018  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 134544)  
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (14-OCT-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
REFERENCE 3 (bases 1 to 134544)  
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (13-DEC-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
COMMENT On Dec 13, 2003 this sequence version replaced gi:39752708.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2202: contig of 2202 bp in length  
\* 2203 2302: gap of unknown length  
\* 2303 134544: contig of 132242 bp in length.  
----- Location/Qualifiers  
1..134544  
/organism="Medicago truncatula"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:3880"  
/clone="mth2-33018"  
/clone\_lib="Medicago truncatula BAC library H2"  
ORIGIN  
Query Match 10.1%; Score 52.6; DB 2; Length 134544;  
Best Local Similarity 45.1%; Pred. No. 37;  
Matches 214; Conservative 0; Mismatches 259; Indels 2; Gaps 1;  
QY 42 ACGAAGTGCCTCCGCGCGGGTACCAGTGCACACCGGGCCAGGATCAGTGGAGTTCGAGCAGCCAGCTC 101  
DB 307 AGGNNNNNNNNNGGG 366  
QY 102 ATCGCGCCTCGCGCGGATCGCGAAACACCGGGCCAGGATCAGTGGAGCAGCAGCGCGATG 161  
DB 367 GGGGCGCGCGGGCGGG 426  
QY 162 CGCGCCTCGATTCGACGCTCGCGCGCAGCTGATCTGTCGCCCGGCTCCACACCCAGG 221  
DB 427 CGCGGGCGG--CCCGCGCGCGCGCGCGGGGGCGCGCGGGCGCGGGGGCGCGCGCGCGCG 484  
QY 222 GGAACCTGCTCGTGGCAGCGCAGCCCGCGCGGGGACAGCTCGTTCAGGCTGGCGTCA 281  
DB 485 CGCGCGCGCGCGCGGGCGGG 544  
QY 282 CGGACCTTCGCGGGTTCGGAGCGGAAACCGCAGCGGCTCAGATTCCTGTCAGTCGATC 341  
DB 545 CGGGGGCGCGCGCGCGGG 604  
QY 342 GGTCTAGTCCGCTGCTCCCTTGCGCTGGAGGATAGCGGTTACAGCAGCAGCGCGACCC 401  
DB 605 GGGCGCGCGCGCGCGGGCGCGGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 664  
QY 402 ACGCGCGCGCGGGTT 461

DB 665 GCGCGCGGGGGGGCGCGCGCGGGCGCGCGCGGGCGCGCGCGGGCGCGCGCGGGCGCGCGCGCG 724  
QY 462 CGGGCGGTCGCGCGAGATCCGTACCGCGCGGACCGCCCTCGCGCGGAGCGCGCGCGGG 516  
DB 725 CGCGCGCGCGCGCGCGCGGGCGCGCGCGGGCGCGCGCGGGCGCGCGCGCGCGCGCGGG 779  
RESULT 39  
LOCUS AE002019/c 12454 bp DNA linear BCT 22-NOV-1999  
DEFINITION Deinococcus radiodurans R1 section 156 of 229 of the complete chromosome 1.  
ACCESSION AE002019 AB0000513  
VERSION AE002019.1 GI:6459547  
KEYWORDS  
SOURCE  
ORGANISM Deinococcus radiodurans R1  
Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
Deinococcaceae; Deinococcus.  
REFERENCE 1 (bases 1 to 12454)  
AUTHORS White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D., Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L., Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M., Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C., Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et al.  
TITLE Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1  
JOURNAL Science 286 (5444), 1571-1577 (1999)  
MEDLINE 20035896  
PUBMED 10567266  
REFERENCE 2 (bases 1 to 12454)  
AUTHORS White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D., Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L., Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M., Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C., Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W., Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S., Smith,H.O., Venter,J.C. and Fraser,C.M.  
TITLE Direct Submission  
JOURNAL Submitted (08-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
source  
1..12454  
/organism="Deinococcus radiodurans R1"  
/mol\_type="genomic DNA"  
/strain="R1"  
/db\_xref="taxon:243230"  
complement(1)  
/gene="DRI172"  
complement(133..540)  
/gene="DRI172"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1334.1"  
/db\_xref="GI:6459548"  
/translation="MSVRGSSSEKVMGRKASAPARESHSSGKATVKEKGRCHPALVAAQADGMTFVAPSAPLPSEAPGTVQRAHRYLLALARQRPDRDYPRAALHWLEDQLIEPCRCAGQLEIVATRGVQAVLGRAAA"  
599..1900  
/gene="DRI173"  
599..1900  
/gene="DRI173"  
/note="similar to GB:M20568 PID:142350 percent identity: 52.60; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI1328.1"  
/db\_xref="GI:6459548"



```
/translation="MLSLGSAHPDARVRAQVALNPNTPAEVLGLLAQFPREVLSNPG
LPRLARGLGSPAEQVTAALLPAPGWNVDSEHEDYGVRTALAAAPGLSPE
RVTALAGAGVQRAQRPDLPEPLPQLAEEDYVRKATARPDPGLDLRTLV
TPDHTLVANARLDLPDCLSLATGDPDLATLARRDLPASVREMLAASAEPL
VRSALQAVPESMLARETADPDVRSALARRDLSPEVLEAHDESGTVRITLL
ENRDPGAILALAAPEPDLQVASEELPAALEALLTDESDVETILAVRDLG
PERLARLAEASAVRGRALAPTLPPELLRQLSADPNAGVRRITLTRELPAPVIRE
LAADPDGVRLLVAGRAGLPAFLRQLAADPDENVREAVG"
/genes="DR1774"
1944..2396
/genes="DR1774"
1944..2396
/genes="DR1774"
1944..2396
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI1341.1"
/db_xref="GI:6459561"
/translation="MALGRQKFLPLADQSTVLNQEDPGRGVSSREGIRFEDFARF
GASKGVESRDLDCQSVSAVGHNVLIIGDAGFOCARLPMKAGQLSPRPHERSV
RSDCSTPRIHRWDQRHAGGSDGAGRTHRSAGSQTEQGRSYPTKSHA"
2469..4706
/genes="DR1775"
2469..4706
/genes="DR1775"
2469..4706
/notes="similar to GB:M87049 Sp:P03018 GB:D00069 GB:X00225
GB:X00738 percent identity: 61.60; identified by sequence
similarity; putative"
/transl_table=11
/product="DNA helicase I1"
/protein_id="AAFI1329.1"
/db_xref="GI:6459549"
/translation="MTSSAGPDLLOALNPTQAAADHFTGPALVIAGSGKTRTLIY
RIAHLGHVHGEILAVFTPKAAEMERAGHLVPGAGDLWMSTFHSAGVRLIT
YGEHIGRGFVYDDDDQDIIKEVMSIPGICAGTQPRVIRGLIDRAKSNLWPTDD
LDRSEFFISGLPRDAAEAYRYEVKKGQNAIDFGDLITETVRLFKVPGVULKVQ
NKAFTHVDYQNTAQRVLELAANKLIENNERDLTKLPVKEAQPVTFHRTDHR
DEGVDVMTLRLHGRGAWSEMAILYRTNAQSVIEESLRVQIPARIVGGVGYDR
REITDLAYARHPADVALRRIIGRPRGIGDTALOKLMEMWARTHTSVLTACAN
AASQNLDRGAHKATFAGLEAMSRADNYEAPALRFVMTSGYLDLROEGQEQ
VRLNELVSAEWSQDANVGGSIADPLDDAALLSSVDVDMRTAKNGAPEDAAT
LMLLNAGKFFPVFVIGVEQGLLESKGAISGSPGIEERELFWVIGITRAMERLLM
TAAQNRMQFTNAASDAFLDIEGLFDVDPYQPIETRAKTWQIYPTPAATIA
VNTSPITASLAVRGGEQVHPKPGEGQVLAVAGVGRQEVTVHFASAGTKLIMVKEA
NLTKL"
complement(4726..5253)
/genes="DR1776"
complement(4726..5253)
/genes="DR1776"
/notes="similar to GP:3191993 percent identity: 45.21;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="MufT/nudix family protein"
/protein_id="AAFI1330.1"
/db_xref="GI:6459550"
/translation="MPFLGVNVGSEWGRVLPGMTDIRLPLGGKLFKFSVRVALLCVRNGE
LLVQTSANAPPEPFFFLPGGALSTDEAATCARREWEETGLLCGLRLVAVVNFPG
PPEKQHEIGFYFLREVEPABLPHGAFSVLNDITDTPFWRVLDLGGARVYPLVRLDL
LSVPVGEVRHIVIRS"
5309..5608
/genes="DR1777"
5309..5608
/genes="DR1777"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI1335.1"
/db_xref="GI:6459555"
/translation="MVPVERQLVITALGLPORTLINGNDYLCQTEGEVVLWLPDPA
POLRVLRALTLPRRVELPCLPDRENAGRFWVPESARVVVVVDSAECAAPAHQCPA"
5810..7105
/genes="DR1778"
5810..7105
/genes="DR1778"
/notes="similar to GB:AB000657 percent identity: 70.88;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="3-isopropylmalate dehydratase, large subunit"
/protein_id="AAFI1331.1"
/db_xref="GI:6459551"
/translation="MGMTYAEKMLAAHSGHDTVPGOLIECATDWLCHETITPAALR
MLEERGMDRVDPQKIVAVPDSHPAMNIAKMYQKLSWQOEKIEHFYDVGOGI
MRVLENTGLMKPGQTLVSGDSTCNAGALGAFATGVSGLDLAGAIYAGKWFKVPET
MLIRVTKTNPVTBKDILEVIKQIGADGANVMMWVGDIYDOLDMEGRPLTNNA
IEAGKGTIVAVDDTTRAYMQORGVSPEQYTESQSDPDAKYKVYVIAIDAQVSPYTAQ
PHIPSGRAGSDRIRVTHAYVSGSCINGRITDLRVARILKGRKVAQDQVQMVVATQ
AIKQAAQEGLEIFVDGASVSPSCAGCIGMHSGVIGPDDICISSNNRNFVRMGD
PSAQIVLASPATVAASAVEGTIADPRYDABGNDLNAAD"
7112..7453
/genes="DR1779"
7112..7453
/genes="DR1779"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI1336.1"
/db_xref="GI:6459556"
/translation="MNPADPPBSWLLTYFVIGLGLLAVRDKWALCGLTTLTAVV
FVLHGPPVFRGLPGLMGLSALGLLALGVSRPLVLAARVGLAALVTLPLTWVVLNAG
FLLFMVAQIS"
7485..8549
/genes="DR1780"
7485..8549
/genes="DR1780"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI1337.1"
/db_xref="GI:6459557"
/translation="MKGDMTLDPMKLLPPGVRPEVTLHPGQLDIRGLDEEAQAVQV
REALKONPLAGVRDRTALVRVGLIRGQALYVAVSETGDAEASWETRPFPDMFG
VPVQDGNAPSTFALHYRGRIDLEPADPDLLDAVEDAVREYHGTEQWABRHAHA
GGPANELPRAVRRLRLEFGLDAGLRVVALPETAAYLLHGGHSEYAPAPQNLADL
APQSEEDDSFFDLFSLFDLHSVTALHAGRVVQGGDLAPALRAALQSHLGADL
GAGDPKWEVALRGLFGLLEGDEVPSGLRVGLLRQWMEADPALAAQAATPSAVTDDG
TWTELDGDDGDEPDEGDE"
8580..9893
/genes="DR1781"
8580..9893
/genes="DR1781"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI1338.1"
/db_xref="GI:6459558"
/translation="MGEAKRKQIQLMPTVPPFAELGRDGEVRVLOGPEDAGORALI
EKALDSQSFQAMDAEYRTVSVLGRGGRGRIATREDVERIPVPAIRQLDGLGALGSA
GOSQAVIFVEGGSVRLREQRHSEGNWQTLPPDLRDPQVLMRALQQHPAFDIBGESL
GQFQADHLLEGRIQVTPDVGELDENGETLEFTEFLVKEFHGQTPESWTAMHREMLEGQ
```

```
Query Match 10.1%; Score 52.4; DB 1; Length 12454;
Best Local Similarity 48.6%; Pred. No. 71;
Matches 143; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 23 GTGCCCCCGCCGCCCATCACGAATGCCACTCCCGCGGGGTGACAGGTGCACAGGTG 82
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12011 GTGCCCCCGCTGCGCGAGGCATCAGGCCAGCCAGCCAGGATCACGGCGCGAGGTC 11952
```



```

Qy 83 GGAGTTGAGCAGCAGTCTATCGCGCTGCGCGGATCCGGAACACCGGCGCAGGAT 142
Db 11951 GGAAGGATGTCGCGGAGAGGTTCTCGGTGAGTCAATCACTGCTAGCGGTGCGGCG 11892
Qy 143 CACGTGAGCAGCAGGATGCGCGCTGAGTCTGAGCTGCGGTGCGCGCAGTCTCGTCCGTC 202
Db 11891 CACGATCAGCATCCCGACGCTGTCCAGTACTCGTGTGTTGAGTGAATCAGCGGTGAGTC 11832
Qy 203 GCCCGGCTCCACACACGAGGAGAACTGCGTGTGCGGAGCGGCGGACGCCCGCGGACAG 262
Db 11831 CGCGTCACGCAACGCTGCGAGTCCGCGCGGSCACAACTCGGACACTTCGAGCAGGTTGCG 11772
Qy 263 CTCGTTCAGGCTGCGGTACGAGACTCTCGGGGTTCGGAGCGGAACGCGGAC 316
Db 11771 CTTGTCCAGCTGTGTCAGCGCGCCCTTCGCGTGTCTCGCGCGGCCCGCAGAGCGGAC 11718

RESULT 40
AC143344
LOCUS Macaca mulatta clone CH250-268P23, DNA linear HTG 10-APR-2003
DEFINITION
AC143344
ACCESSION
VERSION AC143344.1 Gi:29611668
KEYWORDS HTG; HTGS PHASE2; HTGS PGI.
SOURCE Macaca mulatta (rhesus monkey)
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Macaca.
1 (bases 1 to 152686)
Ceuors.M. and Milosavljevic,A.
Pooled genomic indexing (PGI): mathematical analysis and experiment
design
(in) Guigo, R. and Gusfield, D. (Eds.);
ALGORITHMS IN BIOINFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI
2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS: 10-28;
Springer (2002)
2 (bases 1 to 152686)
Milosavljevic,A., Sodergren,E., Ceuors,M., Li,B., Jackson,A.R.,
Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L.,
Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,X., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Egan,A., Earhart,C., Edwards,C.C.,
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
Guaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
Hart,M., Haviak,P., Haves,A., Hernandez,J., Hernandez,O.,
Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homs,F.,
Howard,S., Huber,J., Johnson,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,X., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P.,
Meador,M., Mei,G., Merscher,S., Metzger,M., Miller,A., Miner,G.,
Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogah,M., Okunonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,I., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Roife,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,

```

```

Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
Vinson,R., Wang,O., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Wallington,S., Williams,G., Williamson,A.,
Wleznyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
Zorilla,S., Zucheriapati,R., Weinstein,G. and Gibbs,R.
Direct Submission
Unpublished
3 (bases 1 to 152686)
Worley,K.C.
Direct Submission
Submitted (08-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 152686)
Worley,K.C.
Direct Submission
Submitted (10-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center of Medicine
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: LDMN
Center clone name: CH250-268P23
----- Summary Statistics
Chemistry: Dye-primer Bodypy: Infinity% of reads
Chemistry: Dye-terminator Big Dye: Infinity% of reads
Consensus quality: 3754 bases at least Q40
Consensus quality: 4914 bases at least Q30
Consensus quality: 5932 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: The contigs are based on the application
* of the PGI method using the Human genome (NCBI build 31)
* as the comparative genome.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 152686: contig of 152686 bp in length.
FEATURES
source
1..152686
Location/Qualifiers
/organism="Macaca mulatta"
/mol_type="genomic DNA"
/db_xref="taxon:9544"
/clone="CH250-268P23"
ORIGIN
Query Match 10.0%; Score 52.2; DB 2; Length 152686;
Best Local Similarity 44.2%; Pred. No. 41;
Matches 189; Conservative 0; Mismatches 239; Indels 0; Gaps 0;
Qy 89 GAGCAGCAGCTATCGCGCTGCGCGGATCCGGAACACCGGCGCAGGATCAGTCG 148
Db 56408 GAGCGGGGGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 56467
Qy 149 CAGCAGCGCGGATCGCGCGCTTCGATCTCGAGCGGTGCGCGCGGATCTCGTCG 208
Db 56468 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 56527
Qy 209 CTCCACACACGAGGGAATCGGTCTCGTGTGACGCGCGCGCGCGCGCGCGCGTC 268

```



```

/db_xref="SPTREMBL:Q9FBV4"
/translation="MTDFSPWVAGVVRVLIAPLAGETTMSPVNRLLAARYNVVTGLIS
AMAGPGVGVGRGLGVFEVFLASAVRHQSLASQVSEHLSRALPFWDEGAOKANG
EQPNVFPASGAVFVMLGCRCLTAARTGAACSLARLYDLRSRICIRHQCSLDSLAL
DRVTLVEQVGLAGLPEVIRAHQSLILLRKRKGCENAFVAQAQVAVASWMDVHWRDEV
LWPARLGRVCLADPEGEVAVHAAQVTPPEAVVTVLCRLNRQRVLEDTGQMPHT
LAEPRLILTELARLGRVPLVEQLAASAGALFAWVRACVRRERGAVEDEWAVPA
HEPGLAQLVLELQHGAGNAPVGSLSRAEQAVYVGLAHASVYASBHGHLAVPKYGR
HEGALGANLQNTGUVALPIERAQALHRLDPWNGFPWISWRTHRYRALVHVRKGR
LVDTAGPGLTSLGELWHEQCSKYDDLHVGGQRLADLGRIPAHARSAPRRKSLA
LAFAGLDYAFAAFAVGHGLATSKSTQDGFPLGWLMSQSRARMAEKETDRSRLS
AIDPWNPFPWLAQVYHARKQCNQLLPDGGFAGVGAASAAVLAQCALFEEL
HRCQGLREITGITAQAARQARTVHRTGARIDFAVGLAHARDYINIGHLALPHV
CHNGFPGLRWLASKRGAGAHARTTPAPWCPQMOALALDPWPPWAPAFQWRDVRRLR
LLLAAGLEPPPKLRSWISQTLTQHLLPGQRLQLKOLPV"
6226..7173
/gene="SC00004"
/notes="synonyms: SC1C9.01, SC8E7.40c"
6226..7173
/gene="SC00004"
/notes="SC8E7.40c, unknown, len: 315 aa.
SC1C9.01, hypothetical protein (fragment), len: >64 aa;
identical to C-terminal part of TR:Q9FBV5 (EMBL:AL391338)
Streptomyces coelicolor hypothetical 33.7 kDa protein
SC8E7.40c, 315 aa, situated on the opposite end of the
chromosome"
/codon_start=1
/transl_table=1
/product="hypothetical protein (fragment)"
/protein_id="CAD55410.1"
/db_xref="GI:24413715"
/db_xref="SPTREMBL:Q9BS15"
/translation="MRGIAAAVHPFGMRIRITAFSPFPMTWLSVGEHSHRRVQOA
AAGDGAGIEAAFTVTAAGRLTQLQNDQAASQVEDVAPVSAFVPPGRWPGGWMWS
ATTGRVHVHGAARVTLQWLDRPDVTLGAARVPLVLRDITDGRVSVWVQLFARVA
DGTGLADCPSPPTAGGAGQARWLEAACALVGVAVRLEPPVPPVVAANLRLAGY
RHPNPGPGLRALAEPAGLRPLADGAAVGDPLQVLPVYHALWCGLHTPLDEP
LHEETLVCAERGGGVVEERLGGAGRGGRGRGRPR"
7187..9307
/gene="SC00005"
/notes="SC00005"
/notes="synonym: SC1C9.02"
7187..9307
/gene="SC00005"
/notes="SC1C9.02, possible transposase, len: 706 aa;
identical to TR:Q9FBV6 (EMBL:AL391338), Streptomyces
coelicolor putative transposase SC9E7.39c, 706 aa,
situated on the opposite end of the chromosome. Contains
pfam match to entry PF00665 rve, Integrase core domain"
/codon_start=1
/transl_table=1
/product="putative transposase"
/protein_id="CAD55411.1"
/db_xref="GI:24413716"
/db_xref="GOA:Q9FBV6"
/db_xref="SPTREMBL:Q9FBV6"
/translation="MVEVGAHVARGGTQWVAALQGOVYLLOBDGTETRLLLRLFA
DPGVVGARAPVTPQWLFTVPLAQQRALAWLPHIREVETGWPBPSRGQAM
RPEVDPERTVLAQDAKAKELALGFARVTRTVBRVGRYKQGLGVLDKRAVPA
RGREPTGVADSEVAVLEALRRGRSGKVTGKLVQVIGQILEDTRGVVEMPSPRS
TFYRLSVLADPAPRGPTATAPARASSAPVLRPGVQVQDITRLDIMAILEDG
SLGREPTIADVTRSLAALVLPYSTKAVDAALLAEAWVPHAPPTWPSLHLSR
AEVYPMLESLDERGAARPVVVPETIVVDRGKIYLSQGFVAACMLGVSVQPPAP
RRPOKAVVETFGAINDLFCOHVAGHTGSPQRRLTAAETRTWITPOLQDFDEWI
TCQWNRPHDGLRHPVLPKTLTPNQWAAALITVSGVVPVPLSGADYLELLPVWQPI
TERGIRLDYFVYNDIILDPHRCQSPVASKGKWHVHNPHDAQIIVRLTDQGLHEI
PWIERDHVQFPNEAIWRHVAQVEQGRDQDHEADLADLQLRIRHLAETEQIT
RRRATRGSTAAQPDLPDQRRPFDATAPAPDWSSEDLISVDTAATQGTGSEME
GASVPPAEAGYGLWDAEAEQW"
8231..8623
/gene="SC00005"
/notes="pfam match to entry PF00665 rve, Integrase core
domain, score 47.20, E-value 1.8e-12"
9304..10410

```

```

/gene="SC00006"
/notes="synonyms: SC1C9.03, SCU30.01"
9304..10410
/gene="SC00006"
/notes="SC1C9.03, possible ATP/GTP-binding protein
(fragment), len: >223 aa; identical to N-terminal part of
(fragment), len: >223 aa; DB 1; Length 239050;
Query Match 10.0%; Score 52.2; Pred. No. 34; Mismatches 178; Indels 0; Gaps 0;
Best Local Similarity 47.2%; Conservative 0;
Matches 159;
Qy 131 CCGGGCCAGGATCACGTGCAGCAGCGCGATCGCGCTCTGCACGTCTGCACGTCTGGCGCGCAG 190
Db 205037 CCACTCCATCGCTGGTGCAGCATCGACGTTCGCGCTCGCGCGCGCGTTCGCGTTCAGCAG 205096
Qy 191 CTCGATCTCTGTCGCCCGGCTCCACACACAGGAGAACTGGCTCGTGGGCGAGCGGAGCCCC 250
Db 205097 TACGCGTTCCAGGGCGCTGCCCGCGCGCAGCAGTCTCTCGGCGCGGTTCGTGGCGGTTC 205156
Qy 251 CAGCCGGGACAGCTCGTCCAGGCTGGCGTCAGCGACCTCTCGCGGTCGGGAGCGGAGAAC 310
Db 205157 CAGCAGGGCCACCTGGGCTGCGGTCGAAACGGCGGGCGCGGCGTGGGCGACGGGCGAG 205216
Qy 311 GGCACGGCTCAGATCCCTGTGTCAGTCGATCGCATCGGCTCAGTGCCTGCTCCCTTGGCCCT 370
Db 205217 CGGGCGGAGAAAGTCCCGCATCGGTATGCGGCGCTCAGGTCTCTCGGCGCGAGCGCGGT 205276
Qy 371 GGGAGGATAGCGGTTCACGACGAGCGGCACACGCGGCGCGCGGGCGGGCGGGCGGTTTCAGC 430
Db 205277 CAGCGGGATGCGGTAGGCGGTGACGGTTCGGGGGGCGGTGAGCCAGTGGCGGTTTCGT 205336
Qy 431 CGATCCGCTCGATGACAGCGGCTCGGCGGTTCGGGCG 467
Db 205337 GATACGTCGAGCTGGTGGGTTCGGGTTTCGGGC 205373
RESULT 42
AF272349 2489 bp mRNA linear PRI 21-JAN-2003
LOCUS Homo sapiens quaking protein 3 (QK3) mRNA, partial cds.
DEFINITION AF272349
ACCESSION AF272349
VERSION AF272349.1 GI:17224449
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2489)
Moore, P.L., Jaruzelska, J., Fox, M.S., Urano, J., Firpo, M.T.,
Turek, P.J., Dorfman, D.M. and Reijo Pera, R.A.
Human Pumilio-2 is expressed in embryonic stem cells and germ cells
and interacts with DAZ (Deleted in Azoospermia) and DAZ-Like
proteins
Proc. Natl. Acad. Sci. U.S.A. 100 (2), 538-543 (2003)
JOURNAL PUBMED 12511597
REFERENCE 2 (bases 1 to 2489)
Moore, P.L. and Reijo, R.A.
hQK3 protein interacts with the DAZ protein
Unpublished
JOURNAL
REFERENCE 3 (bases 1 to 2489)
Moore, P.L. and Reijo, R.A.
Direct Submission
TITLE Submitted (26-MAY-2000) OB/GYN/RS, UCSF, Parnassus Ave., San
Francisco, CA 94143, USA
FEATURES
source
1..2489
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
<1..2489
/gene="QK3"
<1..582
/gene="QK3"

```

misc\_feature

gene

/notes="testis-specific; interacts with the DAZ protein, which is involved in male infertility"  
/codon\_start=1  
/product="quaking protein 3"  
/protein\_id="AAL36980.1"  
/db\_xref="GI:17224450"  
/translations="TSPRPFPAAQAALSLDPSERRRLILVRQPPAPART  
RPAETRLPPRGRARFGRERRRLGARRARASACGASARCPGGSAREPE  
REPARGTTPPERPAAGRPRLSLGGGGGGGSELSELYGRNGVEGEAEHPR  
LPDAADQERAEHQALRLDLQPPRAAAGRN"

## ORIGIN

Query Match 10.0%; Score 52; DB 9; Length 2489;  
Best Local Similarity 47.9%; Pred. No. 1.2e+02;  
Matches 181; Conservative 0; Mismatches 195; Indels 2; Gaps 1;  
  
QY 6 AGCAGCAACACCGCGTGTGCGCCCGCGCCCATCAGCAACTGCCACTCCGCGCGGGTG 65  
DB 23 AGCCCGGGCACAGCGCGCGCGCGCTGAGCGGGCTCGACAGCGAGCGCGCGCG 82  
  
QY 66 ACCAGGTGACACAGTGGAGTTGACAGCAGCAGCTATCGCCCTGCGCGCGCATGCG 125  
DB 83 CCGCGCTCTCTCTGTCGCGCGCGCAGAGCGCGCGCGCGCGCGCTCGCGCGCGCG 142  
  
QY 126 AAACACCGCGCGAGATCAGTGCAGCAGCGCGATCGCGCTCTCGATCTCGACGTCGCG 185  
DB 143 AGACCGCTCTCGCGCGCGCGCGCGCTGTGAGGAGCGCGCGCGCGCGCGAG 202  
  
QY 186 CGCAGCTGATCTGTCGCGCGCGCTCCACACAGGGGAACTGCGTCTGTCGCGCGCG 245  
DB 203 AGCGCGCGCGCTGAGGAGCGCG--TCGGGCGCGCGCGGAAAGTCTCTGCGCGCGCG 260  
  
QY 246 AGCCCGACCGCGGACAGCTGTCAGAGCTGCGCTCAGGAGCTCTCGCGCGCTCGGAGCG 305  
DB 261 GAGCGCGCTGTCGCGCGCGCGCGCGCGCTGCGCGCGCGCGAGCGCGCGCGCG 320  
  
QY 306 GAAACGCGCAGCGCTCAGATCCCTGTGTCGATCGCTCAGTCTGTCGCGCGCTGTCGCG 365  
DB 321 GAGCGGACCGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCTCT 380  
  
QY 366 GGCTGGGAGGATAGCGG 383  
DB 381 CTCGCGCGCGCGCGCGG 398

RESULT 43  
AL356119  
LOCUS  
DEFINITION Human DNA sequence from clone RP5-856H7 on chromosome 6, complete sequence.  
ACCESSION AL356119  
VERSION AL356119.31 GI:19351909  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 75152)  
Martin, S.  
Direct Submission  
Submitted (10-MAR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk  
humquery@sanger.ac.uk  
On Mar 11, 2002 this sequence version replaced gi:19309430.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6  
RP5-856H7 is from the library RPI-5 constructed by the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2  
IMPORTANT: This sequence is not the entire insert of clone RP5-856H7. It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
The true right end of clone RP5-856H7 is at 75152 in this sequence.  
The true left end of clone RPI-5LJ12 is at 74160 in this sequence.  
The true right end of clone RPI-495010 is at 2000 in this sequence.  
FEATURES  
Location/Qualifiers  
1..75152  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP5-856H7"  
/clone\_lib="RPI-5"  
23587..23649  
/note="Single clone region. Sequence from reads from a short insert library derived from a single PUC clone. Restriction digest data confirm the assembly."  
42971..43036  
/note="Single clone region. Sequence from reads from a short insert library derived from a single PUC clone. Sequence from overlapping clone RPI-164E20 (AL133542). Assembly confirmed by restriction digest."

## ORIGIN

Query Match 10.0%; Score 52; DB 9; Length 75152;  
Best Local Similarity 47.9%; Pred. No. 52;  
Matches 181; Conservative 0; Mismatches 195; Indels 2; Gaps 1;  
  
QY 6 AGCAGCAACACCGCGTGTGCGCCCGCGCCCATCAGCAACTGCCACTCCGCGCGGGTG 65  
DB 23855 AGCCCGGGCACAGCGCGCGCGCGCTGAGCGGGCTCGACCGCGAGCGCGCGCG 23914  
  
QY 66 ACCAGGTGACACAGTGGAGTTGACAGCAGCAGCTCATCGCGCGCTGCGCGCGCATGCG 125  
DB 23915 CCGGCTCTCTCTGTCGCGCGCGCAGCAGCGCGCGCGCGCGCTGCGCGCGCGCG 23974  
  
QY 126 AAACACCGCGCGAGATCAGTGCAGCAGCGCGATGCGCGCTCGATCTCGACGTCGCG 185  
DB 23975 AGACCGCGCTCTCGCGCGCGCGCGCTGTCGAGGAGCGCGCGCGCGCGCGAG 24034  
  
QY 186 CGCAGCTCAGTCTGTCGCGCGCGCTCCACACAGAGGGAACCTGCTGCTGCTGCGCGCG 245  
DB 24035 AGCGCGCGCGCTGCGAGCGCG--TCGGGCGCGCGCGGAAAGTCTCTCGCGCGCG 24092  
  
QY 246 AGCCCGACCGCGGACAGCTCTGTCAGGCTGCGCTCAGGAGCTCTCGCGCGCTCGGAGCG 305  
DB 24093 GAGCGCGCGTTCGCGCGCGCGCGCGCGCTGCGCGCGGAGCAGAGCGCGCGCGCG 24152  
  
QY 306 GAAACGCGCAGCGCTCAGATCCCTGTGTCAGTCGATTCGCTCAGTTCGCGGTGTCGCGCTT 365  
DB 24153 GAGCGGAGCGCGCGCTCCCGAGCGCGCGCGCGCGCGCTGCGCGCGCGCGCTCTCT 24212  
  
QY 366 GCGCTGGGAGGATAGCGG 383  
DB 24213 CTCGCGCGCGCGCGCGG 24230

```

RESULT 44
AC084804/c
LOCUS
DEFINITION
    219952 bp DNA linear HTG 15-MAY-2002
    Mus musculus clone RP23-314K21 strain C57BL6/J, WORKING DRAFT
    SEQUENCE, 46 unordered pieces.
AC084804
AC084804.10 GI:17488575
VERSION
HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS
SOURCE
    Mus musculus (house mouse)
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
    Grills,G., Han,J., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
    Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
    Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
    High-Throughput Mouse Sequencing
    Unpublished
    2 (bases 1 to 219952)
    Grills,G., Han,J., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
    Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
    Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
    Direct Submission
    Submitted (18-NOV-2000) Department of Molecular Genetics, Albert
    Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
    Bronx, NY 10461, USA
    On Dec 11, 2001 this sequence version replaced gi:13560369.
    -----Genome Center
    Center: Harvard Partners Genome Center
    Center Code: HPGC
    Web site: http://www.hpcg.org/sequence/mouse.html
    Contact: hpgc@mdel.mgh.harvard.edu
    -----Summary Statistics
    Center project name: AEZ
    Sequencing vector: pUC18; L08752
    Chemistry: Dye-terminator Big Dye; 100%
    *Consensus quality: 207914 at least Q20
    *Consensus quality: 205077 at least Q30
    *Consensus quality: 200559 at least Q40
    *Estimated insert size: agarose-FP - N/A
    **Estimated insert size: 219052 - sum-of-contigs
    Quality coverage: agarose-FP - N/A
    Quality coverage: 10.7 x in Q20 bases; sum-of-contigs estimation
    -----
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 46 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    * be preserved.
    * 1 26658: contig of 26658 bp in length
    * 26659 26678: gap of unknown length
    * 26679 53221: contig of 26643 bp in length
    * 53222 53341: gap of unknown length
    * 53342 75821: contig of 22480 bp in length
    * 75822 75841: gap of unknown length
    * 75842 99596: contig of 23755 bp in length
    * 99597 99616: gap of unknown length
    * 99617 120496: contig of 20880 bp in length
    * 120497 120516: gap of unknown length
    * 120517 132931: contig of 12415 bp in length
    * 132932 132952: gap of unknown length
    * 132953 144027: contig of 11076 bp in length
    * 144028 144047: gap of unknown length
    * 144048 153071: contig of 9024 bp in length
    * 153072 153091: gap of unknown length
    * 153092 162146: contig of 9055 bp in length
    * 162147 162166: gap of unknown length
    * 162167 167980: contig of 5814 bp in length
    * 167981 168000: gap of unknown length
    * 168001 173254: contig of 5254 bp in length
    * 173255 173274: gap of unknown length
    * 173275 179246: contig of 5972 bp in length
    * 179247 179266: gap of unknown length
    * 179267 180707: contig of 1441 bp in length
    * 180708 180727: gap of unknown length
    * 180728 184387: contig of 3660 bp in length
    * 184388 184407: gap of unknown length
    * 184408 189237: contig of 4830 bp in length
    * 189238 189257: gap of unknown length
    * 189258 192994: contig of 3737 bp in length
    * 192995 193014: gap of unknown length
    * 193015 197265: contig of 4251 bp in length
    * 197266 197286: gap of unknown length
    * 197287 197589: contig of 304 bp in length
    * 197590 197609: gap of unknown length
    * 197610 199581: contig of 1972 bp in length
    * 199582 199601: gap of unknown length
    * 199602 201006: contig of 1405 bp in length
    * 201007 202394: gap of unknown length
    * 202395 202414: contig of 1368 bp in length
    * 202415 203415: contig of 1001 bp in length
    * 203416 204335: gap of unknown length
    * 204336 204824: contig of 1389 bp in length
    * 204825 204844: gap of unknown length
    * 204845 205380: contig of 536 bp in length
    * 205381 205400: gap of unknown length
    * 205401 207598: contig of 2198 bp in length
    * 207599 207619: gap of unknown length
    * 207620 207875: contig of 257 bp in length
    * 207876 208001: contig of 106 bp in length
    * 208002 208021: gap of unknown length
    * 208022 208140: contig of 119 bp in length
    * 208141 208160: gap of unknown length
    * 208161 208450: contig of 290 bp in length
    * 208451 208470: gap of unknown length
    * 208471 209852: contig of 1382 bp in length
    * 209853 209872: gap of unknown length
    * 209873 210569: contig of 697 bp in length
    * 210570 210589: gap of unknown length
    * 210590 210757: contig of 168 bp in length
    * 210758 210777: gap of unknown length
    * 210778 211360: contig of 583 bp in length
    * 211361 211380: gap of unknown length
    * 211381 211510: contig of 130 bp in length
    * 211511 211530: gap of unknown length
    * 211531 212447: contig of 917 bp in length
    * 212448 212467: gap of unknown length
    * 212468 213049: contig of 582 bp in length
    * 213050 213069: gap of unknown length
    * 213070 213333: contig of 264 bp in length
    * 213334 213353: gap of unknown length
    * 213354 214164: contig of 811 bp in length
    * 214165 214184: gap of unknown length
    * 214185 214400: contig of 216 bp in length
    * 214401 214420: gap of unknown length
    * 214421 214957: contig of 537 bp in length
    * 214958 214977: gap of unknown length
    * 214978 215140: contig of 163 bp in length
    * 215141 215160: gap of unknown length
    * 215161 215345: contig of 185 bp in length
    * 215346 215365: gap of unknown length
    * 215366 216233: contig of 868 bp in length
    * 216234 216253: gap of unknown length
    * 216254 216819: contig of 566 bp in length
    * 216820 216839: gap of unknown length
    * 216840 219659: contig of 2820 bp in length
    * 219660 219679: gap of unknown length
    * 219680 219952: contig of 273 bp in length.
    Location/Qualifiers
  
```

FEATURES

1. 219952 /organism="Mus musculus" /mol\_type="genomic DNA" /strain="CS7BL6/J" /db\_xref="taxon:10090" /clone="RP23-314K21" /sex="male" 1. 26658 /note="assembly\_name:Contig192" 26679. 53321 /note="assembly\_name:Contig191" 5342. 75821 /note="assembly\_name:Contig190" 75842. 99596 /note="assembly\_name:Contig189" 99617. 120496 /note="assembly\_name:Contig188" 120517. 132931 /note="assembly\_name:Contig187" 132952. 144027 /note="assembly\_name:Contig186" 144048. 153071 /note="assembly\_name:Contig185" 153092. 162146 /note="assembly\_name:Contig184" 162167. 167980 /note="assembly\_name:Contig183" 168001. 173254 /note="assembly\_name:Contig182" 173275. 179246 /note="assembly\_name:Contig181" 179267. 180707 /note="assembly\_name:Contig180" 180728. 184387 /note="assembly\_name:Contig179" 184408. 189237 /note="assembly\_name:Contig178" 189258. 192394 /note="assembly\_name:Contig177" clone\_end:SP6 vector\_side:left 193015. 197265 /note="assembly\_name:Contig176" 197286. 197589 /note="assembly\_name:Contig175" 197610. 199581 /note="assembly\_name:Contig174" 199602. 201006 /note="assembly\_name:Contig173" 201027. 202394 /note="assembly\_name:Contig172" clone\_end:T7 vector\_side:right 202415. 203415 /note="assembly\_name:Contig171" 203436. 204824 /note="assembly\_name:Contig170" 204845. 205380 Query Match 10.0%; Score 51.8; DB 11; Length 976; Best Local Similarity 47.8%; Pred. No. 1.6e+02; Matches 236; Conservative 0; Mismatches 253; Indels 5; Gaps 4; QY 25 CGCCCGCGCGCGCCATCAGCACTGCCACTCGCGCGGGGTGACCGAGTTCGACCGGTGG 84 DB 753 CG 694 QY 85 AGTTGAGCAGCGCAGTCTCATCGCGCGCTGCGCGCGCATGCCGAAACACCGCGGCAGATCA 144 DB 693 CG-GGCG 635 QY 145 CGTGAGCAGCGCGAGTGGCGCGCTCGATCTGACAGGTTCGCGCGCA-GTTCGATCTCGTCG 203 DB 634 CG 575 QY 204 CCCTGCTCCACACAGCGGGAACCTGGGTCTGCTGACGCGGAGCGGCGCGCGCGCGCGCG 263 DB 574 CG 515

misc\_feature 1. 219952 /organism="Mus musculus" /mol\_type="genomic DNA" /strain="CS7BL6/J" /db\_xref="taxon:10090" /clone="RP23-314K21" /sex="male" 1. 26658 /note="assembly\_name:Contig192" 26679. 53321 /note="assembly\_name:Contig191" 5342. 75821 /note="assembly\_name:Contig190" 75842. 99596 /note="assembly\_name:Contig189" 99617. 120496 /note="assembly\_name:Contig188" 120517. 132931 /note="assembly\_name:Contig187" 132952. 144027 /note="assembly\_name:Contig186" 144048. 153071 /note="assembly\_name:Contig185" 153092. 162146 /note="assembly\_name:Contig184" 162167. 167980 /note="assembly\_name:Contig183" 168001. 173254 /note="assembly\_name:Contig182" 173275. 179246 /note="assembly\_name:Contig181" 179267. 180707 /note="assembly\_name:Contig180" 180728. 184387 /note="assembly\_name:Contig179" 184408. 189237 /note="assembly\_name:Contig178" 189258. 192394 /note="assembly\_name:Contig177" clone\_end:SP6 vector\_side:left 193015. 197265 /note="assembly\_name:Contig176" 197286. 197589 /note="assembly\_name:Contig175" 197610. 199581 /note="assembly\_name:Contig174" 199602. 201006 /note="assembly\_name:Contig173" 201027. 202394 /note="assembly\_name:Contig172" clone\_end:T7 vector\_side:right 202415. 203415 /note="assembly\_name:Contig171" 203436. 204824 /note="assembly\_name:Contig170" 204845. 205380 Query Match 10.0%; Score 52; DB 2; Length 219952; Best Local Similarity 47.6%; Pred. No. 40; Matches 201; Conservative 0; Mismatches 218; Indels 3; Gaps 2; QY 56 CGCGCGGTGACCGAGTTCGACAGGTGGAGTTGACGAGCGAGTCTATCGCGCGCTCGCG 115 DB 189210 CGTCG 189151 QY 116 CGGATGCGCGAAACACCGCGGCGAGATCAGCTGCGAGCAGCGCGATGCGCGCTCGATCTC 175 DB 189150 CG 189092 QY 176 GACGTCGCGCGCGAGTCTGATCTGTCG--CCCGGCTCCACACAGCGGAACTGGCTC 233



```

QY 304 CGGAACGCGCAGCGGTCTAG--ATCCCTCTTCAGTCCGCTCAGTCCGCTCGTCTCC 361
Db 4476 TGAACGCGGTTCTGTCACGCGCAGCGCCATCCGGGCTTCTGTCGCGGATCGACCC 4417
QY 362 CTTGCGCTGGAGGATAGCGGTTACGACGAGCGGCACACGCGGCGGGCGGGGG 421
Db 4416 GCGGTTCCGGTTCAGACGCGGTCGACGCGGACGCGGCACCGGTCGCGCGCGGTC 4357
QY 422 CGGTTTCAGCGCATCCGCTCGATGACGCGGTCGCGGCTCGGGCGGTCGCGGAGATCC 481
Db 4356 AGGTCGGTTCAGCTGCTGATGACCTGCGACTTCCGCGGACGCGCGGTGATGCC 4297
QY 482 GTACGCGCGGACCGCTCGGCGACGCGCGCG 514
Db 4296 GACGCGCGGCGCGCGCGCGCGCGCGCGAG 4264

RESULT 47
AC010821 208936 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-3K18, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC010821
ACCESSION AC010821.4 GI:9120942
VERSION HTG; HTGS PHASE0.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 208936)
Birren, B., Linton, L., Nusbaum, C. and Lander, B.
Homo sapiens, clone RP11-3K18
Unpublished
2 (bases 1 to 208936)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Becker, R., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Melarim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, K., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul.13, 2000 this sequence version replaced gi:8225265.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTER
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2764
Center clone name: 3_K_18
-----
* NOTE: This record contains 229 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

```

```

* be preserved.
1
697: contig of 697 bp in length
797: gap of 100 bp
1497: contig of 700 bp in length
1597: gap of 100 bp
2265: contig of 668 bp in length
2365: gap of 100 bp
3044: contig of 679 bp in length
3144: gap of 100 bp
3345: contig of 695 bp in length
3939: gap of 100 bp
4642: contig of 703 bp in length
4742: gap of 100 bp
5446: contig of 704 bp in length
5546: gap of 100 bp
6268: contig of 722 bp in length
6368: gap of 100 bp
7066: contig of 698 bp in length
7166: gap of 100 bp
7856: contig of 690 bp in length
7956: gap of 100 bp
8640: contig of 684 bp in length
8740: gap of 100 bp
9415: contig of 675 bp in length
9515: gap of 100 bp
10181: contig of 666 bp in length
10281: gap of 100 bp
10963: contig of 688 bp in length
11063: gap of 100 bp
11760: contig of 691 bp in length
11860: gap of 100 bp
12565: contig of 705 bp in length
12665: gap of 100 bp
13372: contig of 707 bp in length
13472: gap of 100 bp
14162: contig of 690 bp in length
14262: gap of 100 bp
14949: contig of 687 bp in length
15049: gap of 100 bp
15750: contig of 701 bp in length
15850: gap of 100 bp
16524: contig of 674 bp in length
16624: gap of 100 bp
17323: contig of 705 bp in length
17423: gap of 100 bp
18125: contig of 696 bp in length
18225: gap of 100 bp
18911: contig of 686 bp in length
19011: gap of 100 bp
19698: contig of 687 bp in length
19798: gap of 100 bp
20498: contig of 700 bp in length
20598: gap of 100 bp
21313: contig of 715 bp in length
21413: gap of 100 bp
22130: contig of 717 bp in length
22230: gap of 100 bp
22917: contig of 697 bp in length
23017: gap of 100 bp
23713: contig of 696 bp in length
23813: gap of 100 bp
24511: contig of 698 bp in length
24611: gap of 100 bp
25296: contig of 685 bp in length
25396: gap of 100 bp
26074: contig of 678 bp in length
26174: gap of 100 bp
26796: contig of 622 bp in length
26896: gap of 100 bp
27596: contig of 700 bp in length
27696: gap of 100 bp
28400: contig of 704 bp in length
28500: gap of 100 bp

```



\* 28501 29220: contig of 720 bp in length  
\* 29221 29320: gap of 100 bp  
\* 29321 30032: contig of 712 bp in length  
\* 30033 30132: gap of 100 bp  
\* 30133 30825: contig of 693 bp in length  
\* 30826 30925: gap of 100 bp  
\* 30926 31627: contig of 702 bp in length  
\* 31628 31727: gap of 100 bp  
\* 31728 32408: contig of 681 bp in length  
\* 32409 32509: gap of 100 bp  
\* 32509 33192: contig of 684 bp in length  
\* 33193 33292: gap of 100 bp  
\* 33293 33968: contig of 676 bp in length  
\* 33969 34068: gap of 100 bp  
\* 34069 34744: contig of 676 bp in length  
\* 34745 34844: gap of 100 bp  
\* 34845 35226: contig of 682 bp in length  
\* 35227 35627: gap of 100 bp  
\* 35628 36322: contig of 696 bp in length  
\* 36323 36422: gap of 100 bp  
\* 36423 37118: contig of 696 bp in length  
\* 37119 37218: gap of 100 bp  
\* 37219 37946: contig of 728 bp in length  
\* 37947 38046: gap of 100 bp  
\* 38047 38736: contig of 690 bp in length  
\* 38737 38836: gap of 100 bp  
\* 38837 39531: contig of 695 bp in length  
\* 39532 39631: gap of 100 bp  
\* 39632 40118: contig of 687 bp in length  
\* 40119 40418: gap of 100 bp  
\* 40419 41109: contig of 691 bp in length  
\* 41110 41209: gap of 100 bp  
\* 41210 41900: contig of 691 bp in length  
\* 41901 42000: gap of 100 bp  
\* 42001 42619: contig of 619 bp in length  
\* 42620 42719: gap of 100 bp  
\* 42720 43400: contig of 681 bp in length  
\* 43401 43500: gap of 100 bp  
\* 43501 44116: contig of 616 bp in length  
\* 44117 44216: gap of 100 bp  
\* 44217 44916: contig of 700 bp in length  
\* 44917 45016: gap of 100 bp  
\* 45017 45713: contig of 697 bp in length  
\* 45714 45813: gap of 100 bp  
\* 45814 46497: contig of 684 bp in length  
\* 46498 46597: gap of 100 bp  
\* 46598 47293: contig of 696 bp in length  
\* 47294 47393: gap of 100 bp  
\* 47394 48098: contig of 705 bp in length  
\* 48099 48198: gap of 100 bp  
\* 48199 48892: contig of 694 bp in length  
\* 48893 48992: gap of 100 bp  
\* 48993 49666: contig of 674 bp in length  
\* 49667 49766: gap of 100 bp  
\* 49767 50453: contig of 687 bp in length  
\* 50454 50553: gap of 100 bp  
\* 50554 51229: contig of 676 bp in length  
\* 51230 51329: gap of 100 bp  
\* 51330 52017: contig of 688 bp in length  
\* 52018 52117: gap of 100 bp  
\* 52118 52803: contig of 686 bp in length  
\* 52804 52903: gap of 100 bp  
\* 52904 53610: contig of 707 bp in length  
\* 53611 53710: gap of 100 bp  
\* 53711 54406: contig of 696 bp in length  
\* 54407 54506: gap of 100 bp  
\* 54507 55214: contig of 708 bp in length  
\* 55215 55314: gap of 100 bp  
\* 55315 56029: contig of 715 bp in length  
\* 56030 56129: gap of 100 bp  
\* 56130 56834: contig of 705 bp in length  
\* 56835 56934: gap of 100 bp  
\* 56935 57626: contig of 692 bp in length

Query Match 10.0%; Score 51.8; DB 2; Length 208936;  
Best Local Similarity 47.0%; Pred. No. 43;  
Matches 151; Conservative 0; Mismatches 169; Indels 1; Gaps 1;  
  
QY 197 CTCGTGCGCCGGTCCACACACAGGGAACCTGGCTCGTGCAGCGGCGAGCCCGCCAGCCG 256  
Db 100266 CCCCCCGCGGCCCGCCGNNCCNGNGGCCGCGNGGCGGGCGCGCCCGCCCGCCCG 100325  
  
QY 257 GGACAGCTCGTCCAGGCTCGCTCAGCGACCTCTCTCGCGGTCGGGAGCGGAACCGCAC 316  
Db 100326 GNGGCGCCNGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 100385  
  
QY 317 GGCTCAGATCCCTGTCTCAGTCGATCGGTCAGTCGCGTCGCTCCCTTGGCTGGGAGG 376  
Db 100386 CCCCCCGCGCGGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCGCGCGCG 100445  
  
QY 377 ATAGCGGTTCAGACAGAGCGGACACACGCGCGCGCGCGCGCGCGCGCGCGCGCGATC- 435  
Db 100446 CCGGCG 100505  
  
QY 436 CGCTCATGACACAGCGCTGCGGGGTTCGGGGTTCGGCGAGATCGGTACCGCGCGAC 495  
Db 100506 CCGGCG 100565  
  
QY 496 GCCTCG 516  
Db 100566 GNGGCG 100586

## RESULT 48

AL646079 213050 bp DNA linear BCT 02-SEP-2002  
LOCUS Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;  
DEFINITION segment 4/11.  
ACCESSION AL646079 AL646053  
VERSION AL646079.1 GI:17430956  
KEYWORDS  
SOURCE Ralstonia solanacearum  
ORGANISM Ralstonia solanacearum  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Burkholderiaceae; Ralstonia.  
  
REFERENCE 1  
AUTHORS Salanoubat, M., Genin, S., Artiguenave, P., Gouzy, J., Mangenot, S.,  
Arlat, M., Billault, A., Brottier, P., Camus, J.-C., Cattoilco, L.,  
Chandler, M., Choisme, N., Claudel-Renard, C., Cunnac, S., Demange, N.,  
Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T.,  
Sigulier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M.,  
Weissenbach, J. and Boucher, C.A.  
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum  
NATURE 415 (6871), 497-502 (2002)  
MEDLINE 21681879  
PUBMED 11823852  
REFERENCE 2 (bases 1 to 213050)  
AUTHORS Boucher, C.A.  
TITLE Direct Submission  
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,  
F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire  
INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.

## COMMENT

## FEATURES

## source

1. .213050  
/organism="Ralstonia solanacearum"  
/mol\_type="genomic DNA"  
/strain="GMI1000"

```

/db_xref="taxon:305"
/plasmid="megaplasmid"
100..1632
/gene="RSp0487"
/notes="synonym: RS00365"
100..1632
/gene="RSp0487"
/function="macromolecule metabolism; macromolecule
degradation; degradation of proteins, peptides,
glycopeptides"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/protein_id="CADI7638.1"
/db_xref="GI:17430957"
/db_xref="GOA:Q8XS17"
/db_xref="SPTRMBL:Q8XS17"
/translation="MDAPSGRAVIRRAVGRLLLVVGLAVLCIAFNWRHDSR
QLDVPAPLAVDGAARALGEARTISSFTDAQNADQFRLHMLAARYPKAH
AMQRPVGFSLIYTKGSDPSLXPIILMAHQDVVPVAPGTEGWTPPPAGVYKDG
VMRGAWDDKGNLIAQMAAELLAAAGFRPRTHFAFGADEVGGGGAIAALL
KSRGEOLAFVDRGLITGQVPGLTKPAALIGVTKGFLSVALKLSATPGHSSMPA
PGSAAMSAALKHLDDQQLPAGIGVTEMFETLAPKRGINVALSNLWLFPELV
RKQLQSPSTNVLQITTLTVQAGKONVLPGRABATVFNRLLPQDSASSVIAHVE
QAVRSVAGFHEFALPGVSEAAVPSYQTSASYLIGRTVEVPGTVVAPGLMVG
TSRHHVIGISDHFVPSVRPREDIARFHTGNTERISEANVELIRFYHRLHQADA
GA"
complement (1751..3148)
/gene="czcs"
/notes="synonyms: RS00366, RSp0498"
complement (1751..3148)
/gene="czcs"
/EC number="2.7.3.-"
/function="cell processes; protection responses"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE SENSORY TRANSDUCTION
HISTIDINE KINASE FOR COBALT ZINC CADMIUM RESISTANCE"
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CADI7639.1"
/db_xref="GI:17430958"
/db_xref="GOA:Q8XS16"
/db_xref="SPTRMBL:Q8XS16"
/translation="MKRWPSLAARLVAFVLVAGVMLTGAGTYLYALNAQIRATDD
AMLLGIVOLRHCAKDTQSDIDVDRARYTEVVSQHEGLILRVIPGQGRTEANPE
GEVAPAPVPPVYLASEALHTWNSGTGIASRAIAAQAIRGNHRTVTIVVAQGDG
RAALASRHHIPLAVIAGVAVALUSLMLVRSITLKLARIQKQASVPSQDTRLT
LEGAPRSLDLAALAMITRLQDGFARLSQFSADLHFRTPFLANLIGQETLHP
RPAGYRAVLESLEEYARLSRMIEDMLFLARADHARVALSVAFPLDARHEIQVLDYF
EVAADRDTIQVCGNGRVRADITLLRRAINLLDNLARHTPGGARIDVAVOVLPHST
AIAVRNPGPCIIPAASIPLLPDRFYRGDPARANSERSAGLGLAIVQITIMDLHGQTACAS
SPPGGLTEPLIFPA"
complement (3145..3819)
/gene="czcs"
/notes="synonyms: RS00367, RSp0489"
complement (3145..3819)
/gene="czcs"
/function="cell processes; protection responses"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon usage
Predicted by Framed"

```

```

predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC
CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CADI7640.1"
/db_xref="GI:17430959"
/db_xref="GOA:Q8XS15"
/db_xref="SPTRMBL:Q8XS15"
/translation="MRLLIVEDRPKAGDYLLKGLTSGFVADLARSQDGLYOAMEHD
YDLIVLDMPLQMGDQVIBELRRKTPVLFRTADELSDRLKGLLELGDADYIVKPF
AFAPLVARVHTILARGPMRESEVLEADLIDIKRVTAGRIDLTAKAYALLHL
ARTTGVLRSLSISSQVDFVNTNVDVAIRLRARVDDDFEHKLIHLRGMGIV
LDTVAR"
4089..4394
/gene="RSp0490"
/notes="synonym: RS00368"
4089..4394
/gene="RSp0490"
/function="miscellaneous; hypothetical/global homology"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
/protein_id="CADI7641.1"
/db_xref="GI:17430960"
/db_xref="GOA:Q8XS14"
/db_xref="SPTRMBL:Q8XS14"
/translation="MSVRNCLPLAARAAFLACGATASVSEQALSPLOVHCARP
YTDGARGTRDPYTDGGAQAEATLQASVGPDRDPISDGMSSRSPDRDPTDGAQA"
4400..5725
/gene="czcs"
/notes="synonyms: RS00369, RSp0491"
4400..5725
/gene="czcs"
/function="cell processes; protection responses"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE COBALT-ZINC-CADMIUM OUTER MEMBRANE
RESISTANCE PROTEIN"
/protein_id="CADI7642.1"
/db_xref="GI:17430961"
/db_xref="GOA:Q8XS13"
/db_xref="SPTRMBL:Q8XS13"
/translation="MWTHPCIRAVALLSCALLVRSGLAARAPAGAAASDTAAPAQS
APASVPGAHERTLQADIRALATHPALAAGYAVPAGSAGVDAQSLRNPEAFLT
EDVRDRATTAQNLITLELGGKRAARTRAELARDAGQDAVLAALRADVYTRAF
ELSTAQRLRLAQLADALAAQAAVRYQAGKVSPVQTKAAVARAGAAIELRDAQ
ADLSIAQRLASFGGAVAVEADGSAETLPVPPFAHRYLAATAEAPVALASTPRA
LRARLETHRQALVNVRSKRVDPVTLISVGAKDSGANANNAVVGIAVPLPDRNOG
SLLEAQRLAQAEDYBALRLEQAALSQDVARLEAQAQVAGLRQDVLPGAGRAYDA
ARIGFEAGKFNFLDLDLQARTLFGARAQYLAALSRAHQAAAGIDRLIGR"
5752..6966
/gene="czcs"
/notes="synonyms: RS00370, RSp0492"
5752..6966
/gene="czcs"
/function="cell processes; protection responses"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon usage
Predicted by Framed"

```



SWALL: AAL81961 (EMBL:AE010279) (236 aa) fasta scores: E(): 4e-15, 33.04% id in 233 aa  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAE40070.1"  
/db\_xref="GI:33565437"  
/translation="MSIALSNAGAGTKLDEIDLLQWIDAGROVHEADSKRLTMA  
GFVPEGRAGPAAVKLCADLKHSEGLVALDVPAGDVEARRELQSRRAQVAG  
GEVLVSVGVGVMEVFGCTDHTTGFVVVGGVAGGIVAEILLGAPFIRMAFLSARD  
RALRHKAPFIIIDGARGEPADIGAFARLIADVSEFYRRSHLIELDLNPFVVRGRH  
LDPGVVADAGIILQKTF"  
complement(3388..4344)  
/locus\_tag="BPP0662"  
/locus\_tag="BPP0662"  
complement(3388..4344)  
/locus\_tag="BPP0662"  
/note="Similar to Escherichia coli and Escherichia coli  
O157:H7 glycine cleavage system transcriptional activator  
Gcva or B2808 or Ec83668 SWALL:GCVA\_ECOLI  
(SWALL:P32064) (305 aa) fasta scores: E(): 4e-27, 32.33%  
id in 300 aa, and to Rhizobium loti glycine cleavage  
system transcription activator Mlr1973 SWALL:Q98J99  
(EMBL:AP002398) (303 aa) fasta scores: E(): 2.2e-28,  
36.33% id in 300 aa"  
/codon\_start=1  
/transl\_table=11  
/product="LysR-family transcriptional regulator"  
/protein\_id="CAE40071.1"  
/db\_xref="GI:33565438"  
/translation="MQRPPNALRSPESAPRHESFLKAAELHVTGAVSLVKSLE  
SYLVEFMRSRGVAITPGKAPAEIDLEAFNKIGIATDRVQHPGDRGLSICHP  
FAAHLTPWQVQALSTQINVTTLAPELESTDDYDFVVKLGPDPAAQACGGV  
AAERVLVDFTFICSAFQHGELGDIERYELIHAQLRPQDMNWLQSVQOPAL  
ROAPALVPESLITAYNAAGSAGMAIGIHAFIADDIASGRVAVPFPQVRKSHMGFN  
PNAQRAARVPRTEQIRRMVQERNKTRATDTAAALRKSAFAE"  
complement(3445..4083)  
/locus\_tag="BPP0662"  
/note="HMPfam hit to PF03466, LysR substrate binding  
domain, score 3.7e-16"  
complement(4147..4326)  
/locus\_tag="BPP0662"  
/note="HMPfam hit to PF00126, Bacterial regulatory  
helix-turn-helix protein, LysR family, score 6e-23"  
complement(4192..4284)  
/locus\_tag="BPP0662"  
/note="PS00044 Bacterial regulatory proteins, LysR family  
signature."  
complement(4564..5421)  
/locus\_tag="BPP0663"  
complement(4564..5421)  
/locus\_tag="BPP0663"  
/note="Similar to Ralstonia solanacearum putative  
hydroxylase protein hsp0790 or R501917 SWALL:Q8ARPI  
(EMBL:AL646080) (305 aa) fasta scores: E(): 2.3e-16,  
33.61% id in 238 aa, and to Streptomyces coelicolor  
putative hydroxylase Sco4395 or Scd10.27 SWALL:Q9K3V0  
(EMBL:AL359988) (352 aa) fasta scores: E(): 4.5e-15, 31%  
id in 258 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative hydrolase"  
/protein\_id="CAE40072.1"  
/db\_xref="GI:33565439"  
/translation="MQRDYLRLAHGLDFHETWSPQGLPVVMLHGIRGYAETPAGI  
AALQPEVRVIAFDQGRGTDWDCNVYTDVADLAADQSLARFDDLGHSMG  
GINIVYARHGRGELVVEDPAGFSDGSGARIRRELAETPAGFAGVAAADSF  
RSNPSVSEARQORLESMLKPAPEGVYWRVHAGIARLTHLPDPARVVDIAAYVAR  
IGCTLLLRGRSDYIQPAMVNRNALNPRIEATADAGHYIHDDQPAAPAHVCRF  
LKPAPAGA"  
complement(4591..5268)  
/locus\_tag="BPP0663"  
/note="HMPfam hit to PF00561, alpha/beta hydrolase fold,  
score 1e-25"

gene  
complement(5447..6049)  
/locus\_tag="BPP0664"  
complement(5447..6049)  
/locus\_tag="BPP0664"  
/note="Similar to Rhizobium loti hypothetical protein  
Mlr6789 SWALL:Q988D2 (EMBL:AP003010) (196 aa) fasta  
scores: E(): 3e-22, 38.71% id in 186 aa, and to  
Agrobacterium tumefaciens hypothetical protein Atu4237 or  
Aqr\_1.1248 SWALL:AAK89202 (EMBL:AE009353) (200 aa) fasta  
scores: E(): 2.1e-14, 39.08% id in 197 aa"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAE40073.1"  
/db\_xref="GI:33565440"  
/translation="NKTNFSVYHKKWYLOVEDTLAGTGPADGEPLRKIVLAACI  
HNPYAGKYVDLTLDIDSPGAGFRRVRAAQAQOATASVYKCALVADGEYHGN  
ALTNPAANVPDADGAGKSWPSTGKGGPTIIDIPLAHKDALVVRSHYDVSQAQ  
NDGPNDEVLVIAFATRIORLNRALGLAACDVQHDGLN"  
complement(5582..5611)  
/locus\_tag="BPP0664"  
/note="PS00758 ArgE / dapE / ACY1 / CPG2 / yscS family"  
misc\_feature  
Query Match 10.0%; Score 51.8; DB 1; Length 348257;  
Best Local Similarity 48.5%; Pred. No. 38;  
Matches 143; Conservative 0; Mismatches 152; Indels 0; Gaps 0;  
Qy 112 GCACGGCATGCCGAACACCGGCGCAGGATCAGTGCAGCAGCGGATCGCGCTCGA 171  
Db 79205 GCGCGCGCTGACGAGAGCGCTGGCAGCGCTGGGCGGAGCCATCGATCGCGGCTGGC 79264  
Qy 172 TCTCGACGGTGGCGCGCAGCTCGATCTCGTGGCGCGCTCCACACCCAGGGAACTGGC 231  
Db 79265 GCTCGATATCGCGATCGTCCGCGAGCCTTGGCGGGGTTCCGCGGCTGTGTGATCGCGC 79324  
Qy 232 TCGGTGCGACGGCGCCCGCAGCGGAGCAGCTCGTCCAGGCTGGCGTACGACCTCTC 291  
Db 79325 CGACAGCAGCGCGAGCGCGCGCGATCGCGCGCTCGCGCCACGCTCCGCCACGGT 79384  
Qy 292 GCGGTCGGAGCGGAAACCGCAGCGCTCAGATCCTGTCTGTCATCGCATCGGCTCAGTGC 351  
Db 79385 GCGCGTGGCGAGCAGCAGCGCGCGCGCTTCCGCGCGCGCGCGCGCGCGCGTCC 79444  
Qy 352 CGGTGCTCCCTTTGGCTGGAGGATAGCGGTTCACAGCAGCGCGCGCACACCGC 406  
Db 79445 GCGCGCAGCTCGCGCTGGCGGTGTGTGTCTCTCGCGGTGTGTGCGCGTGTCTC 79499  
RESULT 50  
AX805262/c 1293 bp DNA linear PAT 25-NOV-2003  
LOCUS  
DEFINITION Sequence 17 from Patent WO03060128.  
ACCESSION AX805262  
VERSION AX805262.1 GI:38522381  
KEYWORDS uncultured actinomycete  
SOURCE uncultured actinomycete  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
environmental samples.  
REFERENCE  
1  
AUTHORS Farnet, C.M., Staffa, A. and Zazopoulos, E.  
TITLE Compositions, methods and systems for discovery of lipopeptides  
JOURNAL Patent: WO 03060128-A 17 24-JUL-2003;  
Ecopia Biosciences Inc. (CA)  
FEATURES  
source  
1. 1293  
/organism="uncultured actinomycete"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:100235"  
ORIGIN  
Query Match 9.9%; Score 51.6; DB 6; Length 1293;  
Best Local Similarity 48.0%; Pred. No. 1.6e+02;

	Matches	212;	Conservative	0;	Mismatches	224;	Indels	6;	Gaps	2;
Qy	23	GTGCCCCCGCCGCTACGACGACTCCACTCCGGCGGGGTGACACAGGTCCGACCAAGTG	82							
Db	522	GTGCGCTCTGTCCAGCTCGGAGACCCGCTCCGGCGGGTCCGGCCGGCCAGGGC	463							
Qy	83	GGAGTTGAGCAGCCAGCTCATCGCGCTCGCGCGCATGCCGAAACACCGGGCCAGGAT	142							
Db	462	GGCGTAGACCTGCGCCACCCCGGGCGGCTAGACGACTGGCCGGAAGCCGTCCAGCACGGC	403							
Qy	143	CACGTGACGACGCGGATGCGGCTCGATCTGACGCTCGCGCGAGCTCGATCTCGTC	202							
Db	402	GTGCTGTGTAGCCGTGGAACACACGCGCGGCTCGCGCCGAGCAGCAGCGCCCTGGCG	343							
Qy	203	GCCCGGCTCCACACACGAGGGAACCTGGCTCGGTGGCAGCGGCGAGCCCGCCAGCGGACAG	262							
Db	342	GCTCAGCGGGGCCCGGCCCGAGGCGCGACCTCGGCCCATCGCGGCCCCCATCCAGGCGCG	283							
Qy	263	CTCGTCCAGGCTGGGCTACGAGCTCTCGCGGCTCGGGAGCGGAAACGCGCACGG---C	319							
Db	282	GCGCGCCGCTCCGGGTGGGCTCCGCGCGAGGTCCGACGACCTCCAGCGGCTCGGGTC	223							
Qy	320	TCAGATCCCTGTGATCGCATCGGCTCAGTCCCGGTGCTCCCCCTTGGCCTGGGAGGATA	379							
Db	222	GGCGGCCCGGACGAGCTGCCACAGCTGCTCGTCTCGAAGCGCGACCGCAGCGC	163							
Qy	380	GCGGTTACAGCAGCG---GCACACGCGCGGCGGGCGGGCGGGTTTCAGCCGATCC	436							
Db	162	CTCGGCTCTCGGCGCAGCGTGTGCGGACCGCGCGGTTCGAGCAGCCCGGGATCGCGCGCC	103							
Qy	437	GCTCGATGACGAGCGGCTGCGG	458							
Db	102	GATCTCGTAGACGCGCGAG	81							

Search completed: June 27, 2004, 19:39:18  
Job time : 2470.78 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 18:07:12 ; Search time 69.1177 Seconds  
(without alignments)  
4014.538 Million cell updates/sec

Title: US-09-758-759-1\_COPY\_1\_500

Perfect score: 500  
Sequence: 1 ggtaccgcagcgtgcgcgg.....tcgacgtcggtcggaacct 500

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/2/ina/5A\_COMB.seq: \*  
4: /cgn2\_6/prodata/2/ina/5B\_COMB.seq: \*  
5: /cgn2\_6/prodata/2/ina/PCUTUS\_COMB.seq: \*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	149.8	30.0	1401	2	US-08-812-412-1
C 2	149.8	30.0	1401	3	US-09-180-271-4
C 3	56.8	11.4	2122	3	US-09-029-603-1
C 4	50.6	10.1	546	4	US-09-194-905-3
C 5	50.2	10.0	1605	3	US-09-124-541-3
C 6	50.2	10.0	1605	4	US-09-663-326-3
C 7	50.2	10.0	6733	3	US-09-124-541-2
C 8	50.2	10.0	6733	4	US-09-663-326-2
C 9	49	9.8	6854	4	US-09-194-905-7
C 10	48.8	9.8	900	4	US-09-252-991A-5764
C 11	48.8	9.8	2085	4	US-09-252-991A-5708
C 12	48.8	9.8	4776	2	US-08-852-401-1
C 13	48.6	9.7	1416	3	US-08-911-853-3
C 14	48.6	9.7	1416	3	US-09-479-409-3
C 15	48.6	9.7	1416	4	US-09-479-453-3
C 16	48.6	9.7	4377	3	US-08-911-853-28
C 17	48.6	9.7	4377	3	US-09-479-409-28
C 18	48.6	9.7	4377	4	US-09-479-453-28
C 19	48.2	9.6	2082	4	US-08-818-780-67
C 20	48	9.6	3231	1	US-08-074-121-4
C 21	48	9.6	3231	5	PCT-US94-06447-4
C 22	47.8	9.6	1181	4	US-09-635-132-3
C 23	47.8	9.6	1182	1	US-08-341-943-23
C 24	47.8	9.6	2327	6	5229279-5
C 25	47.8	9.6	2327	6	5512669-5
C 26	47.8	9.6	2328	4	US-09-672-749-11
C 27	47.8	9.6	4984	1	US-08-687-806-1

#### ALIGNMENTS

RESULT 1  
US-08-812-412-1/c  
; Sequence 1, Application US/08812412  
; Patent No. 5989869  
; GENERAL INFORMATION:  
; APPLICANT: FILIPPINI, Silvia  
; APPLICANT: LOMOVSKAYA, Natalia  
; APPLICANT: FONSTEIN, Leonid  
; APPLICANT: COLOMBO, Anna L.  
; APPLICANT: HUTCHINSON, C. R.  
; TITLE OF INVENTION: PROCESS FOR PREPARING DAUNORUBICIN AND  
; TITLE OF INVENTION: DOXORUBICIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram  
; STREET: 655 Fifteenth Street N.W. Suite 330  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-5701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/812,412  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitts, Monica C.  
; REGISTRATION NUMBER: 36,105  
; REFERENCE/DOCKET NUMBER: P1615-7005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)638-5000  
; TELEFAX: (202)638-4810  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1401 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Streptomyces peucetius

Sequence 1, Appli  
Sequence 13466, A  
Sequence 13771, A  
Sequence 13657, A  
Sequence 13575, A  
Sequence 9022, Ap  
Sequence 8986, Ap  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 98, Appli  
Sequence 10644, A  
Sequence 10539, A  
Sequence 10409, A  
Sequence 10752, A  
Sequence 9, Appli  
Sequence 76, Appli  
Sequence 122, App  
Sequence 148, App  
Sequence 129, App  
Sequence 1, Appli

US-09-635-132-1  
4 US-09-252-991A-13466  
4 US-09-252-991A-13771  
4 US-09-252-991A-13657  
4 US-09-252-991A-13575  
4 US-09-252-991A-9022  
4 US-09-252-991A-8986  
4 US-09-252-991A-8986  
1 US-07-642-734C-3  
4 US-08-439-009A-3  
2 US-08-804-227C-1  
2 US-08-576-826A-2  
4 US-09-266-965-98  
4 US-09-252-991A-10644  
4 US-09-252-991A-10539  
4 US-09-252-991A-10409  
4 US-09-252-991A-10752  
4 US-09-266-965-9  
4 US-09-266-965-76  
4 US-09-252-991A-122  
4 US-09-252-991A-148  
4 US-09-252-991A-129  
1 US-08-254-357-1

#### RESULT 1

US-08-812-412-1/c  
; Sequence 1, Application US/08812412  
; Patent No. 5989869  
; GENERAL INFORMATION:  
; APPLICANT: FILIPPINI, Silvia  
; APPLICANT: LOMOVSKAYA, Natalia  
; APPLICANT: FONSTEIN, Leonid  
; APPLICANT: COLOMBO, Anna L.  
; APPLICANT: HUTCHINSON, C. R.  
; TITLE OF INVENTION: PROCESS FOR PREPARING DAUNORUBICIN AND  
; TITLE OF INVENTION: DOXORUBICIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram  
; STREET: 655 Fifteenth Street N.W. Suite 330  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-5701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/812,412  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitts, Monica C.  
; REGISTRATION NUMBER: 36,105  
; REFERENCE/DOCKET NUMBER: P1615-7005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)638-5000  
; TELEFAX: (202)638-4810  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1401 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Streptomyces peucetius

C 28 47.8 9.6 4984  
C 29 47 9.4 720  
C 30 47 9.4 1922  
C 31 47 9.4 2175  
C 32 47 9.4 2292  
C 33 46.8 9.4 2571  
C 34 46.8 9.4 3150  
C 35 46.8 9.4 3450  
C 36 46.8 9.4 20235  
C 37 46.8 9.4 43280  
C 38 46.8 9.4 8051  
C 39 46.6 9.3 819  
C 40 46.4 9.3 1029  
C 41 46.4 9.3 1341  
C 42 46.4 9.3 1398  
C 43 46.4 9.3 1446  
C 44 46.4 9.3 3765  
C 45 46.4 9.3 53500  
C 46 46.4 9.2 822  
C 47 46.2 9.2 936  
C 48 46.2 9.2 975  
C 49 46.2 9.2 1431  
C 50 46.2 9.2 1431

```
FEATURE:
; NAME/KEY: CDS
; LOCATION: 190..1398
US-08-812-412-1

Query Match      30.0%; Score 149.8; DB 2; Length 1401;
Best Local Similarity 65.3%; Pred. No. 1.2e-22;
Matches 220; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 164 GGATCCACTGTCGCCAGCGCGGCGGAACTCTCTGCTTCGCGCATGATCTCGTCGGCGT 223
DB 1390 GGATCCAGCGCGCCCTCTGCTGAACTCCGCTCTCGGGGCTGGACTTCGTAGGT 1331
QY 224 GGTTCACGCGAAGAGCAGCGCTAGTCCACCGGCTCGGGCGTGAACCGCTCCGGGGTGC 283
DB 1330 GGTTCACGCGAAGAGCAGCGCTAGTCCGCGTGAAGCGGCGCGAACTCTCGCGCGAGC 1271
QY 284 GCACCGGATGTGTCGCCGGGGTGAAGCGGCTCTGCGCGGGTCTGTCGCACA 343
DB 1270 GGATGGGATGTGGGACCGCGGGAGCCGACGCGCGGATCTTGGCGGCGTGTCTGTAGA 1211
QY 344 CCCAGGAGACGAGTCCGACCGGATCCGAGAAAGTTTCGTCACGGTGGCGCTCTTCGCCG 403
DB 1210 CGCAGCGCAGCATCGGGCCGATCCCGCATAGTTGACACTGTGGCGCTCTTGGCGG 1151
QY 404 TCGCGCGGTACGCCACCAACCGCTTGCCCTCGCGCTTTCGCGAGTTTCAGCAGGCGAGCA 453
DB 1150 TCGCGCGGTACGCCACCAACCGCTTGAGAGCGAGGTCCGCGAGTCTCGCGAGACGCGGCCACCA 1091
QY 464 GGTGGTGGATGCGCTCGACGTCGCGGGCGAACT 500
DB 1090 GGTTCACAGCAGACCGCTTGACCTGGGCGCGCAACTT 1054

RESULT 2
US-09-180-271-4/c
; Sequence 4, Application US/09180271
; Patent No. 6210930
; GENERAL INFORMATION:
; APPLICANT: FILIPPINI, Silvia
; APPLICANT: LOMOVSKAYA, Natalia
; APPLICANT: FONSTEIN, Leonid
; APPLICANT: COLOMBO, Anna L.
; APPLICANT: HUTCHINSON, C. Richard
; APPLICANT: OPTEN, Sharee L.
; APPLICANT: BREME, Umberto
; TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Nikaido, Marmelstein, Murray & Oram
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/180,271
; FILING DATE: 14-JAN-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US98/03938
; FILING DATE: 05-MAR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/901,306
; FILING DATE: 28-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/812,412
; FILING DATE: 06-MAR-1997
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Kitts, Monica C.
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P1615-8010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces peucetius
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 190..1398
US-09-180-271-4

Query Match      30.0%; Score 149.8; DB 3; Length 1401;
Best Local Similarity 65.3%; Pred. No. 1.2e-22;
Matches 220; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 164 GGATCCACTGTCGCCAGCGCGGCGGAACTCTCTGCTTCCTTCCCATGATCTCGTCGGCGT 223
DB 1390 GGATCCAGCGCGCCCTCTGCTGAACTCCGCTCTCGGGCTGGACTTCGTCTAGGT 1331
QY 224 GGTTCACGCGAAGAGCAGCGCTAGTCCACCGCTCGGGCGTGAACCGCTCCGGGGTGC 283
DB 1330 GGTTCACGCGAAGAGCAGCGCTAGTCCGGGTAGGGGGCCCGGAACTCTCTCGCGGAGC 1271
QY 284 GCACCGGATGTGTCGCCGGGGTGAAGCGGCTCTTGGCGGGCTCTTTCGCACA 343
DB 1270 GGATGGGATGTGGGACCGCGGGAGCCGACGCGCGCATTTGGCGGCGTGTGTCTAGA 1211
QY 344 CCCAGGAGACGAGTCCGACCGGATCCGCGAAGTTTCGTACGGTGGCGCTCTTTCGCCG 403
DB 1210 CGCAGCGCAGCATCGGGCCGATCCCGCATAGTTGACACTGTGGCGCTCTTTCGCCG 1151
QY 404 TCGCGCGGTACGCCACCAACCGCTTGCCCTCGCGCTTTCGCGAGTTTCAGCAGGCGAGCA 463
DB 1150 TCGCGCGGTACGCCACCAACCGCTTGAGAGCGAGGTCCGCGAGTCTCGCGAGACGCGGCCACCA 1091
QY 464 GGTGGTGGGATGCGCTCGACGTCGCGGGCGAACT 500
DB 1090 GGTTCACAGCAGACCGCTTGACCTGGGCGCGCAACTT 1054

RESULT 3
US-09-029-603-1/c
; Sequence 1, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Bietenhader, Jurg
; APPLICANT: Toupet, Christine
; APPLICANT: Pospisch, Andreas
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-20555/A/PCT
; CURRENT APPLICATION NUMBER: US/09/029,603
; CURRENT FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: PCT/EP96/03643
; EARLIER FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2122
; TYPE: DNA
; ORGANISM: Streptomyces longisporoflavus
; FEATURE:
; NAME/KEY: misc_RNA
```

; LOCATION: (1)..(2122)  
; OTHER INFORMATION: product = 2.1 kb region  
US-09-029-603-1

Query Match 11.4%; Score 56.8; DB 3; Length 2122;  
Best Local Similarity 66.1%; Pred. No. 0.0014;  
Matches 82; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1 GGTACCGGACCGTGTCCCGGAACAGAGTCCAGATACGGCAGAGAGAACACCCCCCGGT 60  
DB 271 GGGTGGCGTGGCGCCTCGGTGAACGTCTTCTGTGTACGGCAGAGAGAACTTGGCCCTGT 212  
QY 61 AGTCCGGGTAGACGGTGGCGGCGGAAGCGTACGGCTTCGACCGTTCAGCGGGCGGGCGG 120  
DB 211 CGTGGGGAAGACGCGGGGGGTGAACCTCAGCGCTCCCTCGACGCTGAGCGGGCGTGT 152  
QY 121 ACAC 124  
DB 151 TCAC 148

## RESULT 4

US-09-194-905-3/c  
; Sequence 3, Application US/09194905  
; Patent No. 6306627  
; GENERAL INFORMATION:  
; APPLICANT: DECKER, Heinrich  
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR  
; TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS  
; TITLE OF INVENTION: GLA.O AND THEIR USE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/194,905  
; FILING DATE: 29-JUL-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/EP97/02826  
; FILING DATE: 30-MAY-1997  
; PRIOR APPLICATION DATA: DE 19622783.6  
; FILING DATE: 07-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granados, Patricia D.  
; REGISTRATION NUMBER: 33,683  
; REFERENCE/DOCKET NUMBER: 026083/0193  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 546 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-194-905-3

Query Match 10.1%; Score 50.6; DB 4; Length 546;  
Best Local Similarity 49.8%; Pred. No. 0.027;  
Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 239 GCAGCGGTAGTCCACCGCGTCCGGCGTGAACCGTCCGGGTCGCGACCGGATGTGG 298  
DB 383 GAGGAGGAGCGGTACACCTCGTCCGTGGAGACGTGCAAGGGTGCACACCGGTGGCGGA 324  
QY 299 TGGCGGGGTGAGCGCGGCGCTTGGCCCGGCGCTGTGTGTCGACACACCGAGAGACAGGT 358  
DB 323 GCGCGGCGTGCAGCAGGACCTGGGTGCCACGACGCTTGGTGGGGTGAAGCGACCGGTGT 264  
QY 359 CCGACCGATGCGCGCAGAGTTGTCACGCTGGCGCTCTTCGGCGTCCGCGCGGTACGCCA 418  
DB 263 CGGTGATGAGAGCGGTGACGCTGCGACTCGGCCCGCAAGTGCACGATGTCTGTGCGCGC 204  
QY 419 CCACCGCTTGGCCCTCGGCGCTTGAAGCGAGTTCAGCAGGCGGAGCAGGTCCGTCGATGC 478  
DB 203 CGGCCACGCTGTCGACGAGCGCGGTGTCGACACGCTGCCCTGGACGAAGTGAAGCGCG 144  
QY 479 CCTCGACGTCCGGCGCG 495  
DB 143 GATGTCACGACCGCG 127

## RESULT 5

US-09-124-541-3/c  
; Sequence 3, Application US/09124541A  
; Patent No. 6229066  
; GENERAL INFORMATION:  
; APPLICANT: Morris Ph.D., Roy O.  
; TITLE OF INVENTION: A CYTOKININ OXIDASE  
; FILE REFERENCE: UM01490  
; CURRENT APPLICATION NUMBER: US/09/124,541A  
; CURRENT FILING DATE: 1998-07-29  
; EARLIER APPLICATION NUMBER: 60/054,269  
; EARLIER FILING DATE: 1997-07-30  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 1605  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1605)  
US-09-124-541-3

Query Match 10.0%; Score 50.2; DB 3; Length 1605;  
Best Local Similarity 51.0%; Pred. No. 0.033;  
Matches 171; Conservative 0; Mismatches 158; Indels 6; Gaps 2;

QY 165 GATCCACTGTCCGCCAGCCCGCGGAACCTCTCTCTTCGCCATGATCTCGTCCGGGTG 224  
DB 1194 GAGCCACGGGTGCGGCACCCGCCACAGCCCGCTTGTTCGAGCGCCACCTCTCGCGGTG 1135  
QY 225 GTTCAGCGAGAGAGCGCGGTAGTCCACCGGTGCGGGGTGAACCGTCCGGGTCGCG 284  
DB 1134 CACCCGGTCAAGGAACCGCGTAGGCCA---CGTCGCGTGAACCGCAAC---CCCTC 1081  
QY 285 CACCGGATGTGCGTCCCGGGGTGAGCCGCGCTCTTCGCCGCGCTCGTGTGCGCACAC 344  
DB 1080 CAGTAGCTCAGCGTGCACGACGAGCGAGCTCTGTGTCACCGCGCGCGCGCGC 1021  
QY 345 CCAGGAGACAGGTCCGGAACCGATGCGCGAGAACTTCGTACGCTGGCGGTCTTCGCGGT 404  
DB 1020 CGTGGCGTTCGTGCTAGTTGAGCGTGGCTCGATCTGTACACGCTGGTGGCGTTCGCGTC 961  
QY 405 CGCGCGTACGCCACACCGCGCTTCCTCGGCTTCGAGGAGTTCAGAGGCGGAGCAG 464  
DB 960 CCGGCGAGCGCGAGATCCGGCGAGTTCGGCGTTCGGTTCGAGAACCCCGTTCGCGCAG 901  
QY 465 GTCGCTGCGATGCCCTCGAGCTCGGGCGGAACC 499  
DB 900 GTCGCTGCCAGGCTCTGGTTTCAGGAACACCGACC 866



```
RESULT 6
US-09-663-326-3/c
; Sequence 3, Application US/09663326
; Patent No. 6617497
; GENERAL INFORMATION:
; APPLICANT: Morris Ph.D., Roy O.
; TITLE OF INVENTION: A CYTOKININ OXIDASE
; FILE REFERENCE: UMO1490
; CURRENT APPLICATION NUMBER: US/09/663,326
; CURRENT FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 60/054,268
; PRIOR FILING DATE: 1997-07-30
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1605
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1605)
US-09-663-326-3
Query Match      10.0%; Score 50.2; DB 4; Length 1605;
Best Local Similarity 51.0%; Pred. No. 0.033;
Matches 171; Conservative 0; Mismatches 158; Indels 6; Gaps 2;

QY 165 GATCCACTGTCCGACGCGCGGGAACCTCTGCTCTTCCGCAATGATCTCGTCGGCGTG 224
DB 1194 GAGCCACGGGTGCGGCAACCCGCCACAGCCCGAGCTTGTGAGCGCCACCTCTCGCGTG 1135
QY 225 GTTCAGGCGAAGAGCAGCGGTAGTCCACGCGTCGCGGCGTGAACCGTCCGCGGTGCG 284
DB 1134 CACCGGTCAAGAACGCGGTAGGCCA---CGTCGCGTGAACGCGAAC---CCCTC 1081
QY 285 CACCGGATGTGCGTGCAGCGGGGTGAGCGCGCCCTGCTTGCCTGCGCGGTGCTGTCGACAC 344
DB 1080 CACGTAGCTCAGCGTGCAGCGGACGCGAGCTCTGCTGCTCCACCGCGCGCGCGCGC 1021
QY 345 CCAGGAGACCGTCCGACCGATGTCGCGAGAGTTCGTACGCTGGCGCTCTTCGCGGT 404
DB 1020 CGTGGCGTGTGATGATGAGGTGGCTCGATCTGTACACGTTGTTGGGTTCCGCTC 961
QY 405 CGCGCGGTACCGCACCGCCCTGCTTCCCTCGCGCTTGAAGAGTTGAGCGGCGGAGCAG 464
DB 960 CCGGCGAGCGCGAGATCCGCGGCGACGTCGCGCGTGGTGAAGAACCCCGTGTTCGCGCAG 901
QY 465 CTCGCTGCGGATGCGCTCGACGTCCGCGCGGAGCC 499
DB 900 GTCGCTGCGCGGCTCTGTTTACGAACACCGAGCC 866
```

```
RESULT 7
US-09-124-541-2/c
; Sequence 2, Application US/09124541A
; Patent No. 6229066
; GENERAL INFORMATION:
; APPLICANT: Morris Ph.D., Roy O.
; TITLE OF INVENTION: A CYTOKININ OXIDASE
; FILE REFERENCE: UMO1490
; CURRENT APPLICATION NUMBER: US/09/124,541A
; CURRENT FILING DATE: 1998-07-29
; EARLIER APPLICATION NUMBER: 60/054,268
; EARLIER FILING DATE: 1997-07-30
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 6733
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1)..(6733)
```

```
; OTHER INFORMATION: genomic sequence for a cytokinin oxidase from Zea
; OTHER INFORMATION: mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1497)..(2111)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2524)..(3216)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3311)..(3607)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (5697)
US-09-124-541-2
Query Match      10.0%; Score 50.2; DB 3; Length 6733;
Best Local Similarity 51.0%; Pred. No. 0.033;
Matches 171; Conservative 0; Mismatches 158; Indels 6; Gaps 2;

QY 165 GATCCACTGTCCGACGCGCGGGAACCTCTGCTCTTCCGCAATGATCTCGTCGGCGTG 224
DB 3102 GAGCCACGGGTGCGGCAACCCGCCACAGCCCGAGCTTGTGAGCGCCACCTCTCGCGTG 3043
QY 225 GTTCAGGCGAAGAGCAGCGGTAGTCCACGCGTCGCGGCGTGAACCGTCCGCGGTGCG 284
DB 3042 CACCGGTCAAGAACGCGGTAGGCCA---CGTCGCGTGAACGCGAACCC---CTC 2989
QY 285 CACCGGATGTGCGTGCAGCGGGGTGAGCGCGCCCTGCTTGCCTGCGCGGTGCTGTCGACAC 344
DB 2988 CACGTAGCTCAGCGTGCAGCGGACGCGAGCTCTGCTGCTCCACCGCGCGCGCGCGC 2929
QY 345 CCAGGAGACCGTCCGACCGATGTCGCGAGAGTTCGTACGCTGGCGCTCTTCGCGGT 404
DB 2928 CGTGGCGTGTGATGATGAGGTGGCTCGATCTGTACACGTTGTTGGGTTCCGCTC 2869
QY 405 CGCGCGGTACCGCACCGCCCTGCTTCCCTCGCGCTTGAAGAGTTGAGCGGCGGAGCAG 464
DB 2868 CCGGCGAGCGCGAGATCCGCGGCGACGTCGCGCGTGGTGAAGAACCCCGGTTCGCGCAG 2809
QY 465 CTCGCTGCGGATGCGCTCGACGTCCGCGCGGAGCC 499
DB 2808 GTCGCTGCGCGGCTCTGTTTACGAACACCGAGCC 2774
```

```
RESULT 8
US-09-663-326-2/c
; Sequence 2, Application US/09663326
; Patent No. 6617497
; GENERAL INFORMATION:
; APPLICANT: Morris Ph.D., Roy O.
; TITLE OF INVENTION: A CYTOKININ OXIDASE
; FILE REFERENCE: UMO1490
; CURRENT APPLICATION NUMBER: US/09/663,326
; CURRENT FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 60/054,268
; PRIOR FILING DATE: 1997-07-30
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 6733
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1)..(6733)
; OTHER INFORMATION: genomic sequence for a cytokinin oxidase from Zea
; OTHER INFORMATION: mays
; NAME/KEY: CDS
; LOCATION: (1497)..(2111)
; NAME/KEY: CDS
; LOCATION: (2524)..(3216)
; NAME/KEY: CDS
```

LOCATION: (3311)...(3607)  
NAME/KEY: unsure  
LOCATION: (5697)  
US-09-663-326-2

Query Match  
Best Local Similarity 10.0%; Score 50.2; DB 4; Length 6733;  
Matches 171; Conservative 0; Mismatches 158; Indels 6; Gaps 2;

QY 165 GATCCACTGTCGCGGAGCGCGGCGGAGTCTCTGCTTCGCGCATATCTGTCGCGTG 224  
Db 3102 GAGCCACGGTGGCGGACCGCGCACAGCCCCAGCTTGTGAGCGCCACCTCTCCCGTG 3043  
QY 225 GTTCCAGGGAAGAGCAGCGCGTGTGTCACCGCGTGGCGGTGAACCGCGTGGCGGTGG 284  
Db 3042 CACCCGGTCAAGGAACCGCGGTAGGCA--CGTCGCGTGAACGCGAACCC---CTC 2989  
QY 285 CACCGGAGTGTGTCGCGCGGCGGTGAGCGCGCGCTCTTGGCGGCGGTGTCGACAC 344  
Db 2988 CAGTAGCTAGCGTGCACCGCGAGCAGCGAGCTCTGTCACCGCGCGCGCGCGC 2929  
QY 345 CCAGGAGACCGAGTCCGCGACCGATGCGCGAGAGTTCGTACGGTGGCGCTCTTCGCGT 404  
Db 2928 CGTGGCGTGTCTAGTTGAGGTGGCTTCGATGCTGTACAGGTGGTGGCTTCGCTC 2869  
QY 405 CGCGCGTAGCGCACACCGCGTTCGCTCGCGCTTGTGCGGCTTGTGAGCGCGGAGCAG 464  
Db 2868 CCGCGGAGCGCGACGATCGCGCGACGTGCGCGCGTGTGAGAACCCCGTGTTCGCCAG 2809  
QY 465 GTCGCTGCGGATCCCTCGACGTCGCGCGCGAAC 499  
Db 2808 GTCGCTGCCAGGCTCTGTTACGACACCGAC 2774

RESULT 9  
US-09-194-905-7/c  
Sequence 7, Application US/09194905  
Patent No. 6306627  
GENERAL INFORMATION:  
APPLICANT: DECKER, Heinrich  
TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR  
TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS  
TITLE OF INVENTION: GLA.O AND THEIR USE  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESSES:  
ADDRESSES: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/194,905  
FILING DATE: 29-JUL-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/EP97/02826  
FILING DATE: 30-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 19622783.6  
FILING DATE: 07-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Granados, Patricia D.  
REGISTRATION NUMBER: 33,683  
REFERENCE/DOCKET NUMBER: 026083/0193  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6854 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-194-905-7

Query Match  
Best Local Similarity 9.8%; Score 49; DB 4; Length 6854;  
Matches 127; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 239 GCAGCGGTAGTCCACCGCTCGGCGTGAACGGTCCGGGTCGCGACCGGATGTGCG 298  
Db 3734 GAGGGAGCGGTACCTCTGCTCGGTGAGAGTGCAGAGGTGCGCACACCGTGGCGGA 3675  
QY 299 TCCCGGGGTGAGCGCGCTCTCTTGGCGCGGTGCTGTCGCACACCCAGAGACAGGT 358  
Db 3674 CGCGGGGTGCGAGCAGGACCTGGGTGCCACGACGTTGGTGGCGGTGAAGGCAACCGCTGT 3615  
QY 359 CCGACACGATGCCCGCAGAGTTCGTACGTTGGCGCTCTTCGCGTCCGCGTACGCCA 418  
Db 3614 CGGTATGAGCGGTGCGTGCATCGCGCGGAGTGCAGATGCTGTGTCGCGCG 3555  
QY 419 CCACCCGCTTGCCTCGCGCTTGTGAGCGAGTTCAGACGGCGAGCAGTCCGTGCGGATGC 478  
Db 3554 CGGCAGCGTGTGACGAGCGCGTGTGCGCACAGTCCGCTCGACGAGGTGAGGCGCG 3495  
QY 479 CCTCGAGCTCGCGCGCG 495  
Db 3494 GATGTCACGACCGCG 3478

RESULT 10  
US-09-252-991A-5764/c  
Sequence 5764, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 5764  
TYPE: DNA  
LENGTH: 900  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5764

Query Match  
Best Local Similarity 9.8%; Score 48.8; DB 4; Length 900;  
Matches 185; Conservative 0; Mismatches 207; Indels 2; Gaps 1;

QY 105 GTCAGCGGCGGCGGAGACACCGCGGTGAGTCTGCTCCTCTCGCATGATCTCGTGGCGTG 224  
Db 528 GACGCGCGCGTCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 469  
QY 165 GATCCACTGTCGCGCAGCGCGGGAACCTCTCTCTCTCGCATGATCTCGTGGCGTG 224  
Db 468 AGCGCGCAACGGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 411  
QY 225 GTTCCAGGGAAGAGCAGCGGTAGTTCACCGCGTGGCGGTGAAACGCTCCGCGGTGCG 284  
Db 410 ATGCACCTCGCGAGCGCGGTGACGAACCTGTCGACGAAACACCGCGATGTCGCGGTGAC 351  
QY 285 CACCGGATGTGCTGCGCGGCGGTGAGCGCGCGCTGCTGCGCGCGCTGCTGCTGCGCAC 344

Db 350 TTCCTTGTTCGCGTACCGCGGTGTGTACCCGAAACACCGCGGGATCGCAGACAG 291  
QY 345 CCAGGAGACAGTCCGACCGATCGCGCAGAGTTGTACGTCGTCGCTTCGCGGT 404  
Db 290 GCCCATGCCCTGCGCAGCAGCGCGGCCCATCCCGGGCGCATGTGATTTGAC 231  
QY 405 CGCCCGGTACCCACACCCCTTGCCTTCGGCTTCAGGCGAGTTACAGAGGGCGAGCAG 464  
Db 230 GTCGCGGCCCATGCGCGGTGCGAGGAACACACCATCATGATCCCGAGCAGCGTGCGGAGCAG 171  
QY 465 GTCGCTGCGGATGCTTCGAGTCGAGTCGCGGCGGAAC 498  
Db 170 GCCGATGTAGGGCGCGCGCGGATGCGGCTTGAC 137

RESULT 11  
US-09-252-991A-5708  
; Sequence 5708, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5708  
; LENGTH: 2085  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5708

Query Match 9.8%; Score 48.8; DB 4; Length 2085;  
Best Local Similarity 47.0%; Pred. No. 0.064;  
Matches 185; Conservative 0; Mismatches 207; Indels 2; Gaps 1;

QY 105 GTCAGCGGGCGGGCGGACACCGCGGGTCACTGTCAGTGTACGCGGGGAGCTACAG 164  
Db 189 GACGCGCGCGTCTGTCATGTCGTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 248  
QY 165 GATCCACTGTCCGCGACCGCGCGGAACTCTGCTCTTCCGTCATGTCGTCGCGGTG 224  
Db 249 AGCGCGGCAACGGCTGCGCGGTTCGACCATGGGCCACTTCGC--TGAGCTGGCTTTGCC 306  
QY 225 GTTCCAGCGAAGAGCAGCGGTAGTCCACCGCTCGGGGTGTGAACCGCTCCGGGGTGGC 284  
Db 307 ATGCACCTCGCGAGGCGGGTACGAACTCTGTCGACGAACACCGCGCATGTCCGCGTGC 366  
QY 285 CACCGGATGTGCTGCGGGGTGAGCGGCCCTGCTTGGCGCGCTGTCGTGCGACAC 344  
Db 367 TTCCTTGTTCGCGTGACGCGCGGTGTAGCCGAAACACCGGGGATCCGACGACAG 426  
QY 345 CCAGGAGACAGTCCGACCGATGCGCGGAGAGTTGTCACGGTGGCGGTCTTCGCGGT 404  
Db 427 GCCATGCGCTGCGCAGCAGCGCGGCCCATCCCGGGCGATGCGATTGATTTGAC 486  
QY 405 CGCGCGGTACCGACACCCCTTGCCTTCGGCTTTCAGCGAGTTTCAGCAGGGCGAGCAG 464  
Db 487 GTCGCGGCCCATGCGGTGTCGAGGAACACACCATGATCCCGAGCAGCGTGCGGAGCAG 546  
QY 465 GTCGCTGCGGATGCCCTCGAGTCTCGGCGGAAC 498  
Db 547 GCCGATGTAGGGCGCGCGGATGCGGCTTGAC 580

## RESULT 12

US-08-852-401-1/c

; Sequence 1, Application US/08852401  
; Patent No. 5976836  
; GENERAL INFORMATION:  
; APPLICANT: Weber, J. Mark  
; APPLICANT: Hessler, Paul E.  
; APPLICANT: Larsen, Peter E.  
; APPLICANT: Luu, B. Minh  
; TITLE OF INVENTION: Methods and Compositions for Enhancing  
; TITLE OF INVENTION: Erythromycin Production  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rockey, Milnamow & Katz, Ltd.  
; STREET: 2 Prudential Plaza, 180 N. Stetson, Suite  
; STREET: 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852.401  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mueller, Lisa L.  
; REGISTRATION NUMBER: 38,978  
; REFERENCE/DOCKET NUMBER: FER2159POO30US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-616-5400  
; TELEFAX: 312-616-5460  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4776 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-852-401-1

Query Match 9.8%; Score 48.8; DB 2; Length 4776;  
Best Local Similarity 47.4%; Pred. No. 0.064;  
Matches 146; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 192 CTCCTGCTCTTCGCGCATGATCTCGTCGCGGTGTTCCAGGGAAGACAGCGGTAGTC 251  
Db 1644 CACCGCGCGCGCGCATCTCTCTGATCGGTGCGCCACGCTCGTCAGCGCGGTGCA 1585  
QY 252 CACCGCGTCGCGGTGAACCGGTCCGCGGTGCGCACCGGATGTCGTCGCGGGGTGAG 311  
Db 1584 GTAGAGGCCAGCGCATGTCGTGAAGCCGACACCGACAGGTCTCTCGGGGTGCCAG 1525  
QY 312 CCGGCGCTGCTTTCGCGCGGTGTCGTGTCGACACCGCAGAGACCGAGTCCGACCGATGCC 371  
Db 1524 TCCGCGCGCTTTCGCGCATCGCGCATCGCGCGCATGACGTCGAGCAGCAGCATCAC 1465  
QY 372 GCAGAGTTCTGCTGCGGTGGCGCTCTTCGCGGTGCGCGGTGACGACACCGCGTTGCC 431  
Db 1464 GCGGTGGGCGCGGTTCGAGCAGCTTCGCGCATGCGCTTCGCGCGCGCTTCGCGCGC 1405  
QY 432 CTCGCGCTTTCGCGAGTTTCAGCAGGCGGAGCAGGTTCGTCGCGGATGTCCTCGAGTCGCGC 491  
Db 1404 GAACGCGCGGTTCGCGAGCAGCGCGGTTCGCGCGGATGTCGCGCTCTCTCCAGCGCAGC 1345  
QY 492 GCGGAC 499  
Db 1344 GCGCGAGC 1337

## RESULT 13

```

US-08-911-853-3/c
; Sequence 3, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijshbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Genecor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,099
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC3:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-911-853-3

```

Query Match	9.7%;	Score 48.6;	DB 3;	Length 1416;
Best Local Similarity	49.8%;	Pred. No. 0.069;		
Matches 154;	Conservative 0;	Mismatches 149;	Indels 6;	Gaps 1;
Qy	188	GGAACTCCTGCTCCTTCGCCATGATCTCGTCGGCGTGGTTCCAGGCGAAGACGACGGCT	247	
Db	319	GGAGCATCTCGTCGTGATCGAAGGCGTTGGCGATGTATCCACGCGCCCATCTTCATCG		
Qy	248	AGTCCACCGCGTCCGGGCGTGAAACGCTCCGGGGTCGCACCGGAGATGTGGTGC CGGGG	307	
Db	259	AGTCCACCGCGAGCGACAGCTGCATAGTGTGTCATGATCAGTACCGGGTGCCTGCG	200	
Qy	308	TGAGCCGGCCCTGCTTGGCGGGCGTCTGTGTGCAACCCAGGAGACAGAGTCCGACCGGA	367	
Db	199	CCAGCTTGATCAGCTCGGTGCGGGGGCGCGGGCA-----GGCGCAGAGTCGCTGACCA	146	
Qy	368	TGCGCGAGAGTTCGTACAGGTGCGGTCTTCGCGGTGCGCGCTAGCCACACCCGCT	427	
Db	145	CCAGGTGGAAGGTGGAATGCTGTAGCGCTCCTCGGGCTCCTGTGAACCGAACCCGCTCGC	86	
Qy	428	TGCCCTCGGCTTGAAGCGAGTTACAGCAGGGGCGACAGGTGCGTCCGGATGSCCTCGACGT	487	
Db	85	TGACCTGGTACTGGTTCCGTTCCAGCAGGGCGCGACGGCGAGCGGATGATGCTTTCGT	26	
Qy	488	CGCGCGCGA	496	
Db	25	CTTCGACGA	17	

RESULT 14  
US-09-479-409-3/c

```

/ Sequence 3, Application US/09479409
/ Patent No. 6225106
/ GENERAL INFORMATION:
/ APPLICANT: Gerritt, Gijabert
/ APPLICANT: Quax, Wilhelmus J.
/ TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
/ TITLE OF INVENTION: EXPRESSION LEVELS
/ NUMBER OF SEQUENCES: 37
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Genencor International
/ STREET: 925 Page Mill Road
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304-1013
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: Fast-SEQ for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/479,409
/ FILING DATE:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/911,853
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Glaister, Debra J.
/ REGISTRATION NUMBER: 33, 888
/ REFERENCE/DOCKET NUMBER: GC361-2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-846-7620
/ TELEFAX: 650-845-6504
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1416 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-09-479-409-3

```

Query Match	9.7%;	Score 48.6;	DB 3;	Length 1416;
Best Local Similarity	49.8%;	Pred. No. 0.069;		
Matches 154;	Conservative 0;	Mismatches 149;	Indels 6;	Gaps 1;
Qy	188	GGAACTCCTGCTCCTTCGCCATGATCTCGTCGGCGTGTTCAGGCGAAGACGACGCGT	247	
Db	319	GGAGCATCTCGTCTGATCGAAGGCTTGCGCATGTAGTCCA	CGCGCCCATCTTCATCG	260
Qy	248	AGTCCACCGCGTCGGGCGTGAACGCGTCCGGGGTGCACCGGATGTGGGTGCCGGGG	307	
Db	259	AGTCCACCGCCGAGCGCAGCTGCATAGCTGGTCATGATCAGTACCGGGGTGCCGTCGG	200	
Qy	308	TGAGCCGGCCCTGCTTGGCCGGCGTGTGTGCACACCCAGAGACACAGGTCGAGACCGA	367	
Db	199	CCAGCTTGATCAGCTCGGTCCGGGGCGCCGGGCA-----GGCGAGGTTCGTGACCA	146	
Qy	368	TGCCGCAGAGTTCGTACAGTGGCGCTCTTCGCCGTCCGCGCGCTACGCCACCAACCCGCT	427	
Db	145	CCAGGTCGAAGGTCGAATCTGTAGCGCTCTTCGGGCTCTCTGAACCGAACCGGCTCGC	86	
Qy	428	TGCCCTCGGCTTCAGCGAGTTTCAGCAGGCGCAGCAGGTTCGCTCCGATGCTTCGACGT	487	
Db	85	TGACCTGGTACTGTTGCGTTCCAGCAGCGGCGCAGGCGGAGCGATGATGGTTTCGT	26	
Qy	488	CGGCGGCGA	496	
Db	25	CTTCGACGA	17	

RESULT 15  
US-09-479-453-3/c  
; Sequence 3, Application US/09479453

Patent No. 6313283  
GENERAL INFORMATION:  
APPLICANT: Gerritse, Gijbert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
TITLE OF INVENTION: EXPRESSION LEVELS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,453  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/911,853  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J.  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1416 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-479-453-3

Query Match 9.7%; Score 48.6; DB 4; Length 1416;  
Best Local Similarity 49.8%; Pred. No. 0.069;  
Matches 154; Conservative 0; Mismatches 149; Indels 6; Gaps 1;  
QY 188 GGAACTCTGCTCTCCATGATCTCGTCCGCGTGTTCAGCGGAGAGAGAGCGGT 247  
Db 319 GGAGCATCTCGTGTGATCGAAGGCTTGGCGATGTAGTCCACCGCGCCATCTTCATCG 260  
QY 248 AGTCCACCGCTCGGCGTGAACGCTCGGCGTCCGACCGGATGTGCTGCGCGGG 307  
Db 259 AGTCCACCGCGAGCGAGGCTGGCATAGCTGGTGTATGATCAGTACCGGGTGCCTCGG 200  
QY 308 TGAGCGGCGCTTGTGCGCGCTGCTGTGCGCACCCAGGAGACAGGTCCGACCGA 367  
Db 199 CCAGCTTGATCAGCTCGGTGCGCGGGCGCGCGCA-----GGCGAGGTCTGTCACCA 146  
QY 368 TGCCGACAGTCTGTCAGGTGCGCTTTCGCGGTGCGCGCTGCGCACACCGCT 427  
Db 145 CCAGTCAAGGTGCGAATGTGTAGCGCTCTGCGCTCTCTGACACCGACCGCTCGC 86  
QY 428 TGCCCTCGGCTTGAGCGAGTTTCAGCAGGCGAGCAGGTGCGTGGATGCCCTCGAGT 487  
Db 85 TGACCTGTGTTGCTTCCAGCAGCGCGCGAGCGGCGAGCGGATGATGTTTCGT 26  
QY 488 CGGCGCGCA 496  
Db 25 CTTGACGA 17

RESULT 16  
US-09-911-853-28/c  
Sequence 28, Application US/08911853  
Patent No. 6048710

GENERAL INFORMATION:  
APPLICANT: Gerritse, Gijbert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
TITLE OF INVENTION: EXPRESSION LEVELS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/911,853  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/699,092  
FILING DATE: 16-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J.  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-911-853-28

Query Match 9.7%; Score 48.6; DB 3; Length 4377;  
Best Local Similarity 49.8%; Pred. No. 0.071;  
Matches 154; Conservative 0; Mismatches 149; Indels 6; Gaps 1;  
QY 188 GGAACTCTGCTCTCCATGATCTCGTCCGCGTGTTCAGCGGAGAGAGCGGT 247  
Db 1761 GGAGCATCTCGTGTGATCGAAGGCTTGGCGATGTAGTCCACCGCGCCATCTTCATCG 1702  
QY 248 AGTCCACCGCTCGGCGTGAACGCTCGGCGTGGCACCGGATGTGCTGCGCGGG 307  
Db 1701 AGTCCACCGCGAGCGAGGCTGGCATAGCTGTCATGATCAGTACCGGGTGCCTCGG 1642  
QY 308 TGAGCGGCGCTGCTTGGCGCGCTGCTGTGCGCACCCAGGAGACAGGTCCGACCGA 367  
Db 1641 CCAGCTTGATCAGCTCGGTGCGCGGGCGCGCGCA-----GGCGAGGTCTGTCACCA 1588  
QY 368 TGCCGACAGTTCGTCAGGTGCGCTTTCGCGTGGCGCTGCGCGCTACGCCACACCGCT 427  
Db 1587 CCAGTCAAGGTGCGAATGTGTAGCGCTCTGCGCTCTCTGACCGACCGCTCGC 1528  
QY 428 TGCCCTCGGCTTGAGCGAGTTTCAGCAGGCGAGCAGGTGCGTGGATGCCCTCGAGT 487  
Db 1527 TGACCTGTGTTGCTTCCAGCAGCGCGCGAGCGGCGAGCGGATGATGTTTCGT 1468  
QY 488 CGGCGCGCA 496  
Db 1467 CTTGACGA 1459

RESULT 17  
US-09-479-409-28/c  
Sequence 28, Application US/09479409  
Patent No. 6225106  
GENERAL INFORMATION:

APPLICANT: Gerritse, Gijlsbert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
TITLE OF INVENTION: EXPRESSION LEVELS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,409  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/911,853  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-479-409-28

Query Match 9.7%; Score 48.6; DB 3; Length 4377;  
Best Local Similarity 49.8%; Pred. No. 0.071;  
Matches 154; Conservative 0; Mismatches 149; Indels 6; Gaps 1;  
QY 188 GGAACTCTGCTCTTGGCCATGATCTCGTGGCGTGGTTCAGCGGAAGACGCGGT 247  
Db 1761 GGAGCATCTCGTGTGATCGAAGGGCTTGGCGATGTAGTCCACCGGCCCATCTTCATCG 1702  
QY 248 AGTCCACCGCGTGGCGGTGAACGCGTCCGGGTGGCGACCGGATGTGCGTCCGGGG 307  
Db 1701 AGTCCACCGCGGAGCGAGCTGGCATAGCTGTGATCATGATCAGTACCGGGGTGCGTCCG 1642  
QY 308 TGAGCGGCGCTGCTTGGCCGCGTGTGTCGCACACCCAGGAGACCGGTCGCGACCGA 367  
Db 1641 CCAGCTTGATCAGTCTGTCGCGGGGCGCGGCA-----GGCGAGGTGCTGACCA 1588  
QY 368 TGCGCGAGAGTTCGTACGCTGCGCTCTTCCGCGTCCGCGGTACGCCACCCCGCT 427  
Db 1587 CCAGGTGGAAGTTCGGAATGCTGTAGCGCTCTTGGGCGCTCTTGAACCGAACCGGCTCGC 1528  
QY 428 TGCCCTCGGCTTTGAGCGAGTTTCAGCAGGCGGAGCAGGTGCGTCCGATGCCCTCGACGT 487  
Db 1527 TGACCTGGTACTGTTGCGTTCCAGCAGCGGCGCGAGCGGCGGATGATGTTTCGT 1468  
QY 488 CGGCGGCGA 496  
Db 1467 CTTGACGA 1459

RESULT 18  
US-09-479-453-28/c  
; Sequence 28, Application US/09479453  
; Patent No. 633283  
; GENERAL INFORMATION:  
; APPLICANT: Gerritse, Gijlsbert

APPLICANT: Quax, Wilhelmus J.  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
TITLE OF INVENTION: EXPRESSION LEVELS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,453  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/911,853  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-479-453-28

Query Match 9.7%; Score 48.6; DB 4; Length 4377;  
Best Local Similarity 49.8%; Pred. No. 0.071;  
Matches 154; Conservative 0; Mismatches 149; Indels 6; Gaps 1;  
QY 188 GGAACTCTGCTCTTGGCCATGATCTCGTGGCGTGGTTCAGCGGAAGACGCGGT 247  
Db 1761 GGAGCATCTCGTGTGATCGAAGGGCTTGGCGATGTAGTCCACCGGCCCATCTTCATCG 1702  
QY 248 AGTCCACCGCGTGGCGGTGAACGCGTCCGGGTGGCGACCGGATGTGCGTCCGGGG 307  
Db 1701 AGTCCACCGCGGAGCGAGCTGGCATAGCTGTGATCATGATCAGTACCGGGGTGCGTCCG 1642  
QY 308 TGAGCGGCGCTGCTTGGCCGCGTGTGTCGCACACCCAGGAGACCGGTCGCGACCGA 367  
Db 1641 CCAGCTTGATCAGTCTGTCGCGGGGCGCGGCA-----GGCGAGGTGCTGACCA 1588  
QY 368 TGCGCGAGAGTTCGTACGCTGCGCTCTTCCGCGTCCGCGGTACGCCACCCCGCT 427  
Db 1587 CCAGGTGGAAGTTCGGAATGCTGTAGCGCTCTTGGGCGCTCTTGAACCGAACCGGCTCGC 1528  
QY 428 TGCCCTCGGCTTTGAGCGAGTTTCAGCAGGCGGAGCAGGTGCGTCCGATGCCCTCGACGT 487  
Db 1527 TGACCTGGTACTGTTGCGTTCCAGCAGCGGCGCGAGCGGCGGATGATGTTTCGT 1468  
QY 488 CGGCGGCGA 496  
Db 1467 CTTGACGA 1459

RESULT 19  
US-09-818-780-67/c  
; Sequence 67, Application US/09818780  
; Patent No. 667146  
; GENERAL INFORMATION:  
; APPLICANT: McHenry, Charles  
; TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III HOLOENZYME

```
; FILE REFERENCE: 1794_0030004
; CURRENT APPLICATION NUMBER: US/09/818,780
; CURRENT FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/192,736
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 67
; LENGTH: 2082
; TYPE: DNA
; ORGANISM: Thermus thermophilus
; US-09-818-780-67

Query Match          9.6%; Score 48.2; DB 4; Length 2082;
Best Local Similarity 48.1%; Pred. No. 0.084;
Matches 137; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 182 CCGCGGGGAAGTCTGCTGCTTCCGCGATGATCGTCGGGCTGTTCCAGGCGAAGACGA 241
Db 1558 CCAGGAAGTCTCGAGGTCTCCGCTCTTGGCCGCTTCCAGGAGCTCTCCACGTTCT 1499
QY 242 GCGGCTAGTCCACCGCTCGGGGCTGAACGCGTCCGGGGTGGCGACCGGGATGTGGTGC 301
Db 1498 CCAGGCGGTCTCGGCTCTCGGGTAGGCTCCCGGAGGTAGGCGGGTAGTCTGGTGG 1439
QY 302 CCGGGGTGAGCGGCGCTGCTTGGCGGGCTGCTGTGTCGACACCCAGGAGACAGTCCG 361
Db 1438 CCTCGAGGAGTGGCGGAAGAGGCTTCGCGGGCGCGAAGACAGGTCTCTGAAGTCTCT 1379
QY 362 GACCGATGTCGCGAGAAGTCTGTCACGGTGGCGCTCTTCGCGCTGGCGCGCTACGCCACCA 421
Db 1378 CCACAGGCGCACGAAATGCTCTAGGGGCTCGCGGGGGGGAAGTCTTGGCGGCTCTCT 1319
QY 422 CCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 466
Db 1318 TCAGGCTCTCCAGGAGGAAGCCCTTCTCTCTGCGGAGAGT 1274

RESULT 20
US-08-074-121-4/c
; Sequence 4, Application US/08074121
; Patent No. 5767362
; GENERAL INFORMATION:
; APPLICANT: Best, Elaine
; APPLICANT: Knaf, Vic C.
; TITLE OF INVENTION: Methods and Compositions for Modulating
; TITLE OF INVENTION: Lipid Content of Plant Tissues
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: US
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/074,121
; FILING DATE: 08-JUN-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Rae-Venter, Barbara
; REGISTRATION NUMBER: 05938/043001
; REFERENCE/DOCKET NUMBER: 32,750
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-5277
; TELEFAX: (415) 854-0875
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3231 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 861..1328
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1349..2695
; US-08-074-121-4

Query Match          9.6%; Score 48; DB 1; Length 3231;
Best Local Similarity 48.8%; Pred. No. 0.093;
Matches 159; Conservative 0; Mismatches 165; Indels 2; Gaps 1;

QY 47 GAACACCCCCCGGTAGTCCGGGTAGACGGTGGGCGCGAAGCGTACCGGCTTCGACGGT 106
Db 1822 GATGATACACCGGGTAGCGACCTCGCGGGCGATCGCGGGCGGTCTCTTCATCTTCGG 1763
QY 107 CAGCGGGGGCGGGGACACCGG--CGGGTCACTGCTCACGTGTACCGGGGGAGGTACAG 164
Db 1762 CAGCGGGCGGTGCGAGCCCGGACCGTGGGGAACGCGCGGCTTCATGGGCTCTTGGC 1703
QY 165 GATCACAATGTCGCGACGCGCGGAACTCTGCTCTCTGCGCATGATCTCTGCGGGTG 224
Db 1702 CGAAACCTTGTCGCCCATCAGGCGGATCACCTCGCGGGTTCGCGGCGGAGGAGGCC 1643
QY 225 GTTCCAGCGGAGAGCAGCGGTAGTCCACCGCTCGGGCGTGAACCGTCCGGGGTGGC 284
Db 1642 GGAGGTTGATCTGCTGCGGGAAGTGGCGCTTCTCGCGAGGAAGCGGTAGCGGGTG 1583
QY 285 CACCGGAGTGTGTCGCGGGGGTGAAGCGGCGCTCTTGGCGCGGCTGCTGTCGCACAC 344
Db 1582 GATCGCGGTGGCGCGGTGACCTCGCGCGCGCGATGATGATCGCGGGATCTGCAAGTACGA 1523
QY 345 CCAGGAGACCGAGTCCGACCGATGC 370
Db 1522 CTGGGTGGCGGGCGGACCGATGC 1497

RESULT 21
PCT-US94-06447-4/c
; Sequence 4, Application PC/TUS9406447
; GENERAL INFORMATION:
; APPLICANT: Calgene, Inc.
; TITLE OF INVENTION: Methods and Compositions for Modulating
; TITLE OF INVENTION: Lipid Content of Plant Tissues
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weil, Gotshal & Manges
; STREET: 2882 Sand Hill Road, Suite 280
; CITY: Menlo Park
; STATE: California
; COUNTRY: US
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06447
; FILING DATE: 06-JUN-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rae-Venter, Barbara
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: CGNE-097/WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 926-6200
```

TELEFAX: (415) 854-3713  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3231 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 861..1328  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1349..2695  
PCT-US94-06447-4

Query Match 9.6%; Score 48; DB 5; Length 3231;  
Best Local Similarity 48.8%; Pred. No. 0.093;  
Matches 159; Conservative 0; Mismatches 165; Indels 2; Gaps 1;

QY	DB	Score	Indels	Gaps
47	GAACACCCCGGCTAGTCCGGGTAGACGGTGGCGCGGAAAGGGGTACGCGCCCTTCGACGGT	106		
1822	GATGATCACCGGGTAGCCGACCTCGCGGGGATCGCCAGGGGGTCTCTTCATCTTCGG 1763			
107	CAGCGGGGGCGGACACCGG--CGGGTCACTCTGTCACGTGTACGCGGGGACGTACAG 164			
1762	CAGCGGGGGTTCGGAGCCCGGACCGGTGGGGACCGCCGGCGCGCTTCATGCGCTCCTTGGC 1703			
165	GATCCACTGTCCGCCAGCCCGGGAACCTCTCTCTTCGCGCATGATCTCGTGGCGGTG 224			
1702	CGAACTTGTCCGCCATCAGCGGGATCACCTCGGGGGTTCGGCCGACGAGGTGAAGCC 1643			
225	GTTCCAGGGAAGAGACGCGGTAGTCCACCGCGTGGCGGTGAACGCGTCCGGGGTGG 284			
1642	GGAGCGTTCGATCTGTCGCGAAGTCGGCGTTCCTCGCGGAGGAAGCCGTAGCCGGGTG 1583			
285	CACCGGATGTGGTGGCGGGGTGAGCGCGCCCTCTTGGCGGGGTGCTGTGCGACAC 344			
1582	GATCGGGTGGCGCGGTGACCTCGCGCGGGGATGATCGCGGGATTCGAGGTACGA 1523			
345	CCAGGAGACAGTCCGACCGATGC 370			
1522	CTGGGTGGCGGGCGCGACCGATGC 1497			

## RESULT 22

US-09-635-132-3/c  
; Sequence 3, Application US/09635132  
; Patent No. 6620601  
; GENERAL INFORMATION:

APPLICANT: YAMAGUCHI, ISAMU  
APPLICANT: NAKASHITA, HIDEO  
APPLICANT: YOSHIOKA, KEIKO  
APPLICANT: DOI, YOSHIHARU

TITLE OF INVENTION: METHODS FOR TRANSFORMATION OF PLANTS, TRANSFORMED  
TITLE OF INVENTION: PLANTS AND PROCESSES FOR PREPARATION OF POLYESTERS

FILE REFERENCE: 081356/0148

CURRENT APPLICATION NUMBER: US/09/635,132

PRIOR FILING DATE: 2000-08-09

PRIOR APPLICATION NUMBER: JP 11-225832

PRIOR FILING DATE: 1999-08-09

PRIOR APPLICATION NUMBER: JP 11-225839

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 18

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 3

LENGTH: 1181

TYPE: DNA

ORGANISM: Ralstonia eutropha

US-09-635-132-3

Query Match

9.6%; Score 47.8; DB 4; Length 1181;

## RESULT 23

US-08-241-943-23/c  
; Sequence 23, Application US/08241943  
; Patent No. 5602321  
; GENERAL INFORMATION:

APPLICANT: John, Maliyakal E.

TITLE OF INVENTION: TRANSGENIC COTTON PLANTS

TITLE OF INVENTION: PRODUCING HETEROLOGOUS BIOPOLYESTER

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESS: Nicholas J. Seay, Quarles & Brady

STREET: First Wisconsin Plaza, One South

STREET: Pinckney St.,

STREET: P.O. Box 2113

CITY: Madison

STATE: WI

COUNTRY: USA

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/241,943

FILING DATE:

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: us/07/980,521

FILING DATE: 20-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27,386

REFERENCE/DOCKET NUMBER: 11-229-9076-8

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-2484

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 1182 base pairs

TYPE: nucleic acid



```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: NO
; PUBLICATION INFORMATION:
; AUTHORS: PEOPLES
; JOURNAL: J. Biol. Chem.
; VOLUME: 264
; PAGES: 15293-15297
; DATE: 1989
; US-08-241-943-23

Query Match
Best Local Similarity 46.8%; Score 47.8; DB 1; Length 1182;
Matches 185; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 105 GTCACGGGGGGGGGACACCGGGGTCAGCTGTCAGCTGTCAGCGGGGACGTACAG 164
DB 411 GGCATCGCCCATGCGGAACCATCGCGAGCCCGGACAGACGTGGGGGGGGGCGCTCAT 352
QY 165 GATCCACTGTCCGCGAGCCCGGGGAACCTCTGCTCTCTTCGCCATGATCTGTCGGGT- 223
DB 351 GTTTTCCTGGCGCGGCCACACAGCATCTCGCGGTGCGCCGCCATGATCGGTTGGCGGC 292
QY 224 --GGTTCAGGGAAGACAGCGCGGTAGTCCACCGCGTCGGCGTGAACCGGTCCGGGGT 281
DB 291 CAGCATCAGCGCTTCAGCGCCGAGCCGACACCTTGTGATGTCATGCGCGCACCAT 232
QY 282 GCGCACCGGATGTCGTGCGGGGGTGAGCGCGCCCTGCTTGGCGCGCTGCTGTGCGCA 341
DB 231 CCGCGGACGGCGGCTTGATCGCGGCTGCGGTGCGGGTCTGCGCCGGAACCGCGGT 172
QY 342 CACCCAGAGACAGCTCGGACCGCATCGCGGAGTTCGTCAGCGGTGCGCTTTCGC 401
DB 171 CAGCACCTGCGCCATGATGACTTCGCTCACTGCTCGGCTTGAGCGCGCGGCTCCAG 112
QY 402 CCGCGCGGTGAGCGACCGACCGGCTGCGCTCGGCTTGAGCGAGTTCAGCAGGGCGAG 461
DB 111 CCGGCGCTTGATGACCAACCGGACCGGACCGGTCGGTCCGGTCCGGGATCTTGGCCAGGCGGCC 52
QY 462 CAGGTGCGGTGCGGATGCGCTCGACGTGCGCGCGCA 496
DB 51 AAACCTGCGGACCGCGGTGCGGCGCGGATACGA 17

RESULT 24
5229279-5/c
; Patent No. 5229279
; APPLICANT: PEOPLES, OLIVER P.; SINSKEY, ANTHONY J.
; TITLE OF INVENTION: METHOD FOR PRODUCING NOVEL POLYESTER
; BIOPOLYMERS
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/556,535
; FILING DATE: 13-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 67,695
; FILING DATE: 29-AUG-1987
; SEQ ID NO:5:
; LENGTH: 2327
5229279-5

Query Match
Best Local Similarity 46.8%; Score 47.8; DB 6; Length 2327;
Matches 185; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 105 GTCACGGGGGGGGGACACCGGGTCAAGTCTGCTCGCCATGATCTCGTTCGGCGGT- 223
DB 450 GGCATCGCCCATGCGGAACCATCGCGCGAGCCCGGACGACGTCGCGGGGCGGCTCAT 391
QY 165 GATCCACTGTGCGCGAGCGCGGGGAACCTCTGCTCTCTTCGCCATGATCTCGTTCGGCGGT- 223
DB 390 GTTTTCCTGGCGCGGCCACACAGCATCTCGGCGTTCGCCGCTGATGATCGCGGCG 331
QY 224 --GGTTCAGGGAAGACAGCGGTGTCACCGGTGCGGGGTCGAAACGCTTCGGGGT 281
DB 330 CAGCATCAGCGCTTCAGGCGCGGACGACACCTTGTGATGTCATGCGCGCGCACCAT 271
QY 282 GCGCACCGGATGTCGTGCGGGGTGAGCGCGGCTCTGCTTGGCGCGGTCTGTGTCGCA 341
DB 270 CGCGCGGAGCGCGGCTTGTATCGCGGCTGCGGTGCGGGTTCGCGCCGGAACCGCGGT 211
QY 342 CACCCAGGAGACAGGTTCGGACCGGATGCGCGAGAGTTCGTCAGGTGCGGTCTTCGC 401
DB 210 CAGCACCTGCGCCATGATGACTTCGCTCACTGCTCGGCTTTCGACGCGCGGCGGCTCCAG 151
QY 402 CGTCGCGCGGTGAGCGACCGGCTGCGCTTGGCGCTTTCGAGCGAGTTCAGCAGGGCGAG 461
DB 150 CGCGGCGCTTGTATGATGACCGGACCGGACCGGATCTTGGCCAGGCGCGGCC 91

RESULT 25
5512669-5/c
; Patent No. 5512669
; APPLICANT: PEOPLES, OLIVER P.; SINSKEY, ANTHONY J.
; TITLE OF INVENTION: GENE ENCODING BACTERIAL ACETOACETYL-COA
; REDUCTASE
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,667
; FILING DATE: 29-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 124,570
; FILING DATE: 20-SEP-1993
; APPLICATION NUMBER: 944,488
; FILING DATE: 03-NOV-1992
; APPLICATION NUMBER: 566,535
; FILING DATE: 13-AUG-1990
; APPLICATION NUMBER: 67,695
; FILING DATE: 29-JUN-1987
; SEQ ID NO:5:
; LENGTH: 2327
5512669-5

Query Match
Best Local Similarity 46.8%; Score 47.8; DB 6; Length 2327;
Matches 185; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 105 GTCACGGGGGGGGGACACCGGGTCAAGTCTGCTCGCCATGATCTCGTTCGGCGGT- 164
DB 450 GGCATCGCCCATGCGGAACCATCGCGCGAGCCCGGACGACGTCGCGGGGCGGCTCAT 391
QY 165 GATCCACTGTGCGCGAGCGCGGGGAACCTCTGCTCTCTTCGCCATGATCTCGTTCGGCGGT- 223
DB 390 GTTTTCCTGGCGCGGCCACACAGCATCTCGGCGTTCGCCGCTGATGATCGCGGCG 331
QY 224 --GGTTCAGGGAAGACAGCGGTGTCACCGGTGCGGGGTCGAAACGCTTCGGGGT 281
DB 330 CAGCATCAGCGCTTCAGGCGCGGACGACACCTTGTGATGTCATGCGCGCGCACCAT 271
QY 282 GCGCACCGGATGTCGTGCGGGGTGAGCGCGGCTCTGCTTGGCGCGGTCTGTGTCGCA 341
DB 270 CGCGCGGAGCGCGGCTTGTATCGCGGCTGCGGTGCGGGTTCGCGCCGGAACCGCGGT 211
QY 342 CACCCAGGAGACAGGTTCGGACCGGATGCGCGAGAGTTCGTCAGGTGCGGTCTTCGC 401
DB 210 CAGCACCTGCGCCATGATGACTTCGCTCACTGCTCGGCTTTCGACGCGCGGCGGCTCCAG 151
QY 402 CGTCGCGCGGTGAGCGACCGGCTGCGCTTGGCGCTTTCGAGCGAGTTCAGCAGGGCGAG 461
DB 150 CGCGGCGCTTGTATGATGACCGGACCGGACCGGATCTTGGCCAGGCGCGGCC 91
```

```
QY 462 CAGGTGCGTGGGATGCGGCTCGACGTGGCGGCGGA 496
Db 90 AAACCTGCCGACCGCGGTGCGGCGCGGATACGA 56

RESULT 26
US-09-672-749-11/c
; Sequence 11, Application US/09672749
; Patent No. 6475734
; GENERAL INFORMATION:
; APPLICANT: Liebergessel, Matthias
; APPLICANT: Fallis, Patricia L
; APPLICANT: Dong, Jian G
; APPLICANT: Li, Chun Ping
; APPLICANT: Nichols, Scott E
; TITLE OF INVENTION: POLYHYDROXYALKANOATE SYNTHASE GENES
; FILE REFERENCE: 5718-120
; CURRENT APPLICATION NUMBER: US/09/672,749
; CURRENT FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/156,770
; PRIOR FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2328
; TYPE: DNA
; ORGANISM: Ralstonia eutropha
US-09-672-749-11

Query Match 9.6%; Score 47.8; DB 4; Length 2328;
Best Local Similarity 46.8%; Pred. No. 0.1;
Matches 185; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 105 GTGAGCGGCGGCGGACACCGCGGTCAGTGTACGCGGGGACGTACAG 164
Db 450 GGCATCGCCATGCGGAACCATCGCGGAGCCCGGACGACGTGCGGGCGGCTCAT 391
QY 165 GATCCACTGTGCGGACGCGGCGGGAATCTCTGCTTTCGCGCATGATCTGCGGGGT- 223
Db 390 GTTTTCTGCGCGCGCGGACACGATCTCGCGGTGCGCGGCTGATCGGCTTGGCGGC 331
QY 224 --GTTTCCAGGCGAAGACGACGCGGTAGTCCACCGGTCGCGGCGTGAACGCGCTCGGGGT 281
Db 330 CAGCATCAGGCTTCAGCGCCGAGCGGACACCTTGTGATGTCATGCGCGGACCAT 271
QY 282 GCGCACCGGAGTGTGCGTGCAGCGGCGGTGAGCGGCGGCTTGTGCGGCGGTGCGTGTGCGCA 341
Db 270 CGCGGCGAGCGGCGCTTGTGATCGCGGCTGCGGCTTGTGCGGCGGCTTGTGCGGCGGCT 211
QY 342 CAGCCAGGAGACGAGTCCGAGCCGATGCGGAGGATTCGTCAGGTCGCGGCTTGTGCG 401
Db 210 CAGCACTGCGCATGATGATCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 151
QY 402 CGTGGCGGCTAGCGCACCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 461
Db 150 CGCGGCTTGTGATCAGCGCACCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 91
QY 462 CAGGTGCGTGGGATGCGGCTCGACGTGGCGGCGGA 496
Db 90 AAACCTGCCGACCGCGGTGCGGCGCGGATACGA 56

RESULT 27
US-08-687-806-1/c
; Sequence 1, Application US/08687806
; Patent No. 5811272
; GENERAL INFORMATION:
; APPLICANT: Snell, Kristi D.
; APPLICANT: Hogan, Scott A.
; APPLICANT: Sim, Sang Jun
; APPLICANT: Sinekey, Anthony J.
; APPLICANT: Rha, Chokyun

; TITLE OF INVENTION: Method for Controlling Molecular Weight of
; TITLE OF INVENTION: Polyhydroxyalkanoates
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,806
; FILING DATE: 11-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: pabst, patrea l.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: mit 6867
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-687-806-1

Query Match 9.6%; Score 47.8; DB 1; Length 4984;
Best Local Similarity 46.8%; Pred. No. 0.1;
Matches 185; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 105 GTGAGCGGCGGCGGACACCGCGGTCAGTGTACGCGGGGACGTACAG 164
Db 3106 GGCATCGCCATGCGGAACCATCGCGGAGCCCGGACGACGTCGCGGGCGGCGCTCAT 3047
QY 165 GATCCACTGTGCGGACGCGGCGGTAGTCCACCGGTCGCGGCGGTGTAACGCGTCCGCGGT 223
Db 3046 GTTTTCTGCGCGCGGCGGACCATCTCGGGGTGCGCGGCGGCTGATCGGTTGGCGGC 2987
QY 224 --GTTTCCAGGCGAAGACGAGCGGTCAGTCCACCGGTCGCGGCGGTGTAACGCGTCCGCGGT 281
Db 2986 CAGCATCAGGCGCTTTCAGGCGCGGACGCGGTCGCGGCGGTGATGCTGATGCTGATGCTGATG 2927
QY 282 GCGCACCGGAGTGTGCGTGCAGCGGCGGTGAGCGGCGGCTGCTTGGCGCGGCTGCTGCTGCTG 341
Db 2926 CGCGGCGAGCGCGGCTTGTGATCGCGGCTGCGGCGGCTTGTGCGGCGGCTTGTGCGGCGGCT 2867
QY 342 CAGCCAGGAGACGAGTCCGAGCCGATGCGGCGGAGAGTTCGTCAGGTGCGGCTTGTGCG 401
Db 2866 CAGCACCTGGCGCATGATGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2807
QY 402 CGTGGCGGCTAGCGCACCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 461
Db 2806 CGCGGCGTGTGATGACCGCGGACCGGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 2747
QY 462 CAGGTGCGTGGGATGCGGCTCGACGTGGCGGCGGA 496
Db 2746 AAACCTGCCGACCGCGGTGCGGCGCGGATACGA 2712

RESULT 28
US-09-635-132-1/c
; Sequence 1, Application US/09635132
; Patent No. 6620601
; GENERAL INFORMATION:
```

APPLICANT: YAMAGUCHI, ISAMU  
APPLICANT: NAKASHITA, HIDEO  
APPLICANT: YOSHIOKA, KEIKO  
APPLICANT: DOI, YOSHIHARU  
TITLE OF INVENTION: METHODS FOR TRANSFORMATION OF PLANTS, TRANSFORMED  
TITLE OF INVENTION: PLANTS AND PROCESSES FOR PREPARATION OF POLYESTERS  
FILE REFERENCE: 081356/0148  
CURRENT APPLICATION NUMBER: US/09/635,132  
PRIOR FILING DATE: 2000-08-09  
PRIOR APPLICATION NUMBER: JP 11-225832  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: JP 11-225839  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4984  
TYPE: DNA  
ORGANISM: Ralstonia eutropha  
US-09-635-132-1

Query Match 9.6%; Score 47.8; DB 4; Length 4984;  
Best Local Similarity 46.8%; Pred. No. 0.1;  
Matches 185; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 105 GTCAGCGGGGGGGGACACCGCGGGTCACTGTCACGTGTACGCGGGGACGTACAG 164  
Db 3106 GGCATCGCCCATGCGGAAACCATCGCGGAGCGCGGACGACGTCGTGGGGCGGCTCAT 3047  
QY 165 GATCCACTGTCGCGAGCGCGGGAACCTCTGCTCTTCGCCATGATCTCGTCGGGT - 223  
Db 3046 GTTTCCTGGCGCGCGGCCACCATGTCGGGTGCGCGCCGATGATCGGTGGCGGC 2387  
QY 224 --GGTTCAGCGAAGACGAGCGGTATGACCGCGTTCGCGGGGTGAACCGTCCGGGGT 281  
Db 2986 CAGCATCACGCGCTTCAGCGCGGAGCGGACACCTTGTGATGTCATGCGCGCACCAT 2927  
QY 282 CGGCACCGGATGTCGTCGCGGGGTGAGCGGCGCTGTCGCGCGGTGTCGCGCA 341  
Db 2926 CGCGGAGCGCGGCTTGATCGCGGCTGCGGCGGTTCGCGCGGCTTCGCGCGGAAACGCGGT 2867  
QY 342 CACCCAGGAGACCGGTTCGCGGACGATGCGCGGAGTTGTCACGGTGGCGCTCTTCGC 401  
Db 2866 CAGCACCTGGCCATGATGACTTCGCTCACCTGTCGCGGTGACGCGCGCGCTCCAG 2807  
QY 402 GTCGCGCGGTACGCCACACCGCTTGCCTCGGCTTGAGCGAGTTCAGCAGGCGAG 461  
Db 2806 CGCGGCTTGATGACCCAGGACCGGATTCGCGGTGCGGGATCTTGCCACGAGCGGCC 2747  
QY 462 CAGGTGCGGTGCGGATGCGCTCGAGTTCGCGCGCGA 496  
Db 2746 AAACCTGCGGCGGCGGTGCGGCGCGGATACGA 2712

RESULT 29  
US-09-252-991A-13466/c  
Sequence 13466, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 13466  
LENGTH: 720  
TYPE: DNA

Query Match 9.4%; Score 47; DB 4; Length 1992;  
Best Local Similarity 45.3%; Pred. No. 0.15;  
Matches 170; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 114 CGGGCGGACACCGCGGGTCACTGTCACGTGTACGCGGGGACGTACAGGATCCACTG 173  
Db 1231 CGCAGGACACCGTCTGAGGACGATCGCCAGGGCCACGCGCTGGCGAGGATCCACAGGT 1290  
QY 174 TCCGCCAGCGCGGCGGAACTCTGCTCTTCCCATGATCTCTGCGCGGTGTCAGGC 233  
Db 1291 GTCGAACGCGCGTTGGGACACCGGCTCGTAGCTGATGGCGAAGTCCGCGAGGC 1350  
QY 234 GAAGACGAGCGGTAGTCCACCGCGTTCGCGGTGAGAACCGGTCCGGGTCCGACCGGAT 293  
Db 1351 GACCGCAGGACGTTGGCCACCAACCGCCGCGGCTGACCTCGCGCTAGGAACGCGAG 1410

ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13466

Query Match 9.4%; Score 47; DB 4; Length 720;  
Best Local Similarity 45.3%; Pred. No. 0.15;  
Matches 170; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 114 CGGGCGGACACCGCGGGTCACTGTCACGTGTACGCGGGGACGTACAGGATCCACTG 173  
Db 638 CGCAGGACACCGTCTGAGGACGATCGCCAGGGCCACGCGCTGGCGAGGATCCACAGGT 579  
QY 174 TCCGCCAGCGCGGCGGAACTCTGCTCTTCCCATGATCTCTGCGCGGTGTCAGGC 233  
Db 578 GTCGAACGCGCGTTGGGACACCGCGGTCTAGCATCTGCGAAGTCCGCGAGGC 519  
QY 234 GAAGACGAGCGGTAGTCCACCGGTCTGCGGTGCGGCTGAAACGCTCCGGGTGCGCACCGGAT 293  
Db 518 GACCGCAGGAGCTTGGCCACCAACGCGCGCGGTGACCTCGCGTAGGAACGCCAG 459  
QY 294 GTGCGTCCGGGGGTGAGCGGCGCTGTTGGCGGGGTCTGTCGCGACACCCAGGAGAC 353  
Db 458 CGGCTTGAGCGGCGCGAAGAACCACTGTTCCGGCAAGCGCTGCGCTAGTCGCGACCCG 399  
QY 354 CAGGTCCCGGACCGATGCGCGAGAGTTCTGTCACGTGCGCGCTCTTCGCGCTCGCGCCGTA 413  
Db 398 CCGCTCCCGCGGTAGCGGACTTTGGCGAAGCCGCGTGCCTGTACAGGCTTCCAG 339  
QY 414 CGCCACACCGCGTTGCCCTCGGCTTACGAGAGTTACAGCGGCGAGCAGGTCGGTCCG 473  
Db 338 CTCGCGCAGCGCATCTCTGCTTCCCGCTGCTCTGCGGTACCAACCACTCGCGCTG 279  
QY 474 GATGCCCTCGAGTC 488  
Db 278 CTCGCGCTGTTGCC 264

RESULT 30  
US-09-252-991A-13771  
Sequence 13771, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 13771  
LENGTH: 1992  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13771

Query Match 9.4%; Score 47; DB 4; Length 1992;  
Best Local Similarity 45.3%; Pred. No. 0.15;  
Matches 170; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 114 CGGGCGGACACCGCGGGTCACTGTCACGTGTACGCGGGGACGTACAGGATCCACTG 173  
Db 1231 CGCAGGACACCGTCTGAGGACGATCGCCAGGGCCACGCGCTGGCGAGGATCCACAGGT 1290  
QY 174 TCCGCCAGCGCGGCGGAACTCTGCTCTTCCCATGATCTCTGCGCGGTGTCAGGC 233  
Db 1291 GTCGAACGCGCGTTGGGACACCGGCTCGTAGCTGATGGCGAAGTCCGCGAGGC 1350  
QY 234 GAAGACGAGCGGTAGTCCACCGCGTTCGCGGTGAGAACCGGTCCGGGTCCGACCGGAT 293  
Db 1351 GACCGCAGGACGTTGGCCACCAACCGCCGCGGCTGACCTCGCGCTAGGAACGCGAG 1410

QY 294 GTGCGTCCGCGGGGTGAGCGCGCCCTGCTTGGCCGCGTGTGTGCACACCCAGGAGAC 353  
Db 1411 CGCCTTGGCGGCGCCGAGACACAGTGTTCGGCGACGCGCTGGGTGTGCTGCGACCGC 1470  
QY 354 CAGGTCCGACCGATCGCGAGAGTTCGACGGTGGCGCTTTCGCGTCCGCGCGTA 413  
Db 1471 CCGTCCGCGCGTACGCGCACTTGGCAAGACCGCGCTGTACAGGCTTCCAG 1530  
QY 414 CGCCACACCGCTTCCCTTCGCGCTTACGAGTTTCAGACGGCGAGCAGTCCGTGCG 473  
Db 1531 CTGCGCCAGCGGATCTCTGGCTTCGCGCGTCTCTCGGTACAGCACTTCGCGTG 1590  
QY 474 GATGCCCTCGACGTC 488  
Db 1591 CTCGCGCTGTTGCC 1605

RESULT 31  
US-09-252-991A-13657/c  
; Sequence 13657, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13657  
; LENGTH: 2175  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13657

Query Match 9.4%; Score 47; DB 4; Length 2175;  
Best Local Similarity 45.3%; Pred. No. 0.15;  
Matches 170; Conservative 0; Mismatches 205; Indels 0; Gaps 0;  
QY 114 CGGCGGACACCGCGGGTCACTCGTCACTGTCAGCGGGGAGTACAGGATCCACTG 173  
Db 675 CGCAGGACACCGTCCGAGGACGATCCGAGGCGCCGCTGGCGAGGATCCAGGGT 616  
QY 174 TCGCGCAGCGCGGAACTCTGCTTCCGCAATGATCTCGTGGCGGTGTTCCAGGC 233  
Db 615 GTCGAACGCGCGGTTCGGGACCAACGCGGTCTGATGCGGAAACAGTCCGAGGC 556  
QY 234 GAAGAGCAGCGGTAGTCCACCGCGTCGCGGCGTGAACGCGTCGCGGGTCCGACCGGGAT 293  
Db 555 GACCGCCAGGAGCTTGGCCACCAACGCGCGCGGTGACCTCGCGTAGGAACGCAAG 496  
QY 294 GTGCGTCCGCGGGTGAAGCGCGCTTGGCCGCGTGTGTCGACACCCAGGAGAC 353  
Db 495 CCGCTTGAAGCGGCGGCAAGAACAGTGTTCGGCAAGGCGTGGCGTAGTCCGCGCGG 436  
QY 354 CAGGTCCGACCGATCGCGCAGAGTTCGTCAGGTGGCGCTTTCGCGTCCGCGCGTA 413  
Db 435 CCGTCCGCGGTAGCGGACCTTGGCAAGACCGCGCTGCGCTGTACAGGCTTCCAG 376  
QY 414 CGCCACACCGCTTCCCTTCGCGCTTGAAGGAGTTTCAGAGGGGAGCAGTCCGTGCG 473  
Db 375 CTGCGCCAGCGGATCTCTGGTCTCCGCGCTCTCGGTACAGCACTTCGCGGTG 316  
QY 474 GATGCCCTCGACGTC 488  
Db 315 CTCGCGCTGTTGCC 301

RESULT 32  
US-09-252-991A-13575/c  
; Sequence 13575, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13575  
; LENGTH: 2292  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13575

Query Match 9.4%; Score 47; DB 4; Length 2292;  
Best Local Similarity 45.3%; Pred. No. 0.15; Indels 0; Gaps 0;  
Matches 170; Conservative 0; Mismatches 205; Indels 0; Gaps 0;  
QY 114 CGGCGGACACCGCGGGTCACTCGTCACTGTCAGCGGGGAGTACAGGATCCACTG 173  
Db 646 CCGCAGGACACCGTCCGAGGACGATCCGCGGCGCGCGTCCGAGGATCCACAGGGT 587  
QY 174 TCGCGCAGCGCGGGAATCTCTCTTCGCCATGATCTCGTCCGCGTGTTCAGGC 233  
Db 586 GTCGAACGCGCGTTCGGGACCAACGCGGTCTGATGCTGATGCGGAACGATGCGCGAGC 527  
QY 234 GAAGAGCAGCGGTAGTCCACCGCGTCGCGGTGAACGCGTCGCGGGTCCGACCGGGAT 293  
Db 526 GACCGCCAGGAGCTTGGCCACCAACGCGCGGCGTACCTCGCGTAGGAACCGCAGAG 467  
QY 294 GTGCGTCCGCGGGTGAAGCGCGCTCTTGGCGGCGTCTGTGCGCACACCCAGGAGAC 353  
Db 466 CCGCTTGAAGCGGCGGCAAGAACACAGTGTTCGGGCAAGCGCTGCGCGAGCCTCG 407  
QY 354 CAGGTCCGACCGATCGCGCAGAGTTCGTCAGGTGGCGCTCTTCGCGCGTCCGCGCGTA 413  
Db 406 CCGTCCGCGGTAGCGGCACTTGGCGAAGACCGCGTCCGCTGTACAGCGTTCAG 347  
QY 414 CGCCACACCGCTTGGCGCTTCGCGCTTGAAGCGGATTCAGCAGGCGGAGGTCGCTGCG 473  
Db 346 CTCGCGCAGCGGCACTCTCTGGCTTCGCGCGTCTGTGCGGTACCAAGCACTTCGCGGTG 287  
QY 474 GATGCCCTCGACGTC 488  
Db 286 CTCGCGCTGTTGCC 272

RESULT 33  
US-09-252-991A-9022  
; Sequence 9022, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9022  
; LENGTH: 2571

; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9022

Query Match 9.4%; Score 46.8; DB 4; Length 2571;  
Best Local Similarity 47.6%; Pred. No. 0.16;  
Matches 138; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 208 ATGATCTCTCGCGGTGTCAGGCGAAGACGCGGTAGTCCACCGGTGGCGGTG 267  
Db 971 ATGCTGCGAGGAGTAGGGGATGGCGATCGGTGTGCGCATCTTGTGCGCGAGCTG 1030  
QY 268 AACCGTCCGGGGTCCGACCGGATGTCGTGCGGGGTGAGCGGCTCTCTTGGCC 327  
Db 1031 GAGCTTTCTCGCGGAGACGCCGATGACCTGGACGCGACGTTCTCGATGAGGTAG 1090  
QY 328 GCGTCTGTGTCACACCCAGGAGACCGATGTCGAGCCGATGCCGAGAAAGTTGCTACG 387  
Db 1091 CGGCCGATCGGGTTGGGCTGCCGACAGCTTCTTGCGCACCTTGTAGCCGATCACC 1150  
QY 388 GTGCGCTCTTCGCGGTGCGCGCTAGCCACACCGCTTGCCCTGGGCTTGAGCGAG 447  
Db 1151 ACCGTGTGGGGCTCTTGTGCGCTCGGTGAGTAGTGTGCTCGGACCGCCAG 1210  
QY 448 TTCAGCAGGCGACGAGTGGTGGGATGCCCTCGACGTGGCGCGAA 497  
Db 1211 TTGAGGATTTCCGGGAAGTCGTGTTGTCGCGCGACGTAGCGTGGTA 1260

RESULT 34  
US-09-252-991A-8765/c  
; Sequence 8765; Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8765  
; LENGTH: 3150  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8765

Query Match 9.4%; Score 46.8; DB 4; Length 3150;  
Best Local Similarity 47.6%; Pred. No. 0.16; Mismatches 152; Indels 0; Gaps 0;  
Matches 138; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 208 ATGATCTCTCGCGGTGTCAGGCGAAGACGCGGTAGTCCACCGGTGGCGGTG 267  
Db 2588 ATGCTGCGAGGAGTAGGGATGCGATGCGGTGTGCGCATCTTGTGCGCGAGCTG 2529  
QY 268 AACCGTCCGGGGTCCGACCGGATGTCGTGCGGGGTGAGCGGCTCTCTTGGCC 327  
Db 2528 GAGCCTTTCTCGCGGACGACCGCCGATGACCTGGAACGCGACGTTCTCGATGAGGATGTAG 2469  
QY 328 GCGTCTGTGTCACACCCAGGAGACCGAGTCCGAGCCGATGCCGAGAAAGTTCTGTCACG 387  
Db 2468 CGGCCGATCGGGTTGGGCTGCCGACAGCTTCTTGGCGACCTTGTAGCCGATCACC 2409  
QY 388 GTGCGCTCTTCGCGGTGGGCGCTAGCCACACCGCTTGCCCTCGGCTTGGAGCGAG 447  
Db 2408 ACCGTGTGGGGCTCTTCTGCGCTCGTGAAGTAGTGTGCGCTCGGACCGGCGAG 2349  
QY 448 TTCAGCAGGCGAGCAGTGGTGGGATGCCCTCGACGTGGCGCGAA 497

Db 2348 TTGAGGATTTCCGGGAAGTCGTGTTGTCGCCGCGAGTAGGCGTGTA 2299

RESULT 35  
US-09-252-991A-8986/c  
; Sequence 8986; Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8986  
; LENGTH: 3450  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8986

Query Match 9.4%; Score 46.8; DB 4; Length 3450;  
Best Local Similarity 47.6%; Pred. No. 0.16;  
Matches 138; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 208 ATGATCTCTCGCGGTGTCAGGCGAAGACGCGGTAGTCCACCGGTGGCGGTG 267  
Db 1458 ATGCTGCGAGCGGATAGGGATGCGGTGTGCGCATCTTGTGCGCGAGCTG 1399  
QY 268 AACCGTCCGGGGTCCGACCGGATGTCGTGCGGGGTGAGCGGCTCTCTTGGCC 327  
Db 1398 GAGCCTTTCTCGCGGAGACCGCGATGACCTGGAACGCGACGTTCTCGATGAGGTAG 1339  
QY 328 GCGTCTGTGTCACACCCAGGAGACCGATGTCGAGCCGATGCCGAGAAAGTTGCTACG 387  
Db 1338 CGGCCGATCGGGTTGGGCTGCCGACAGCTTCTTGGCGACCTTGTAGCCGATCACC 1279  
QY 388 GTGCGCTCTTCTCGCGGTGCGCGGTAGCCACCGCTTGGCTTGGCTTGGAGCGAG 447  
Db 1278 ACCGTGTGGGGCTCTTCTGTCGCGCTGCGGTGAGTAGTGTGCTCGGACCGGCGAG 1219  
QY 448 TTCAGCAGGCGAGCAGTGGTGGGATGCCCTCGACGTGGCGCGAA 497  
Db 1218 TTGAGGATTTCCGGGAAGTCGTGTTGTCGCCGCGACGTAGGCGTGTA 1169

RESULT 36  
US-07-642-734C-3/c  
; Sequence 3; Application US/07642734C  
; Patent No. 5824513  
; GENERAL INFORMATION:  
; APPLICANT: Katz, L  
; APPLICANT: Donadio, S  
; APPLICANT: McAlpine, J B  
; TITLE OF INVENTION: Recombinant DNA Method for Producing  
; TITLE OF INVENTION: Erythromycin Analogs  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Edward H. Gorman  
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
; CITY: Park Rd  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6"
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"

```



OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6"  
OTHER INFORMATION: 6-deoxyerythronolide B formatio"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10723..15165  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10831..12174  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl:ACP synthase domain of modul"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 12379..13350  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 14062..14610  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 14857..15114  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15166..20235  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15172..16569  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl:ACP synthase domain of modul"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 16768..17721  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 18379..18921  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19149..19398  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19492..20235  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: thioesterase domain of module 6"  
US-08-439-009A-3

Query Match 9.4%; Score 46.8; DB 3; Length 20235;  
Best Local Similarity 46.4%; Pred.No.0.17; Indels 0; Gaps 0;  
Matches 153; Conservative 0; Mismatches 177;  
QY 109 GCGGGGGGGGACACCGGGGTACGTCGTCACGTGTACGCGGGGACGTACAGGATC 168  
DB 12569 GCGGGGTGACGACGTCGACCGCGTCCAGCGGTGCGCGCGGTGAGCAGGTGCGTCAGC 12510  
QY 169 CACTGTCCCGCCCGCGGGAACCTCTGTCTTCCCATGATCTCGTCGCGGTGTTTC 228  
DB 12509 GACCACTGACGTGCGGGGGCCAGCGCCCGCTCGACGTCCGGATCGAGTCGCGCAATACC 12450  
QY 229 CAGCGGAAGACGCGGTAGTCCACCGCGTGGGGTGAAAGCGTCCGGGGTGGCGACC 288

DB 12449 TCGGACTCCCGCAGCAGGTGCGGGGGCCATCCCTGCACTGCGCGCCCTGCGCGGGGAC 12390  
QY 289 GGGATGTGCTGCGCGGGGGTGAAGCCGCCCTGCTTGGCCGCGCTGCTGTCGACACCCGAG 348  
DB 12389 ACCATCGCGACGCGCGCGGAGTTTCGGGCTTGGCCGCTGCGCGGTGCGCGATGCGG 12330  
QY 349 GAGACCAAGTCCGACACCGATGCCGACAGAGTTGTCACGGTGGGCTCTTTCGCGGTGCGG 408  
DB 12329 CCGTCGGCCACCGCGCGGAGCCCGGCGAGGCGTCTGCGGTGCGCGACCAACCGCG 12270  
QY 409 CCGTACGCGCACACCGCGCTTGCCTTCGCGCC 438  
DB 12269 CGGTGCTCCACGCGCTGCGCGGGTGGCC 12240

RESULT 38  
US-08-804-227C-1/C  
; Sequence 1, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; CITY: LILLY CORPORATE CENTER  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 43280 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 816..14234  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14351..19945  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20010..31199  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 31232..36067  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 36249..41774  
; US-08-804-227C-1

Query Match

9.4%; Score 46.8; DB 2; Length 43280;



Best Local Similarity 50.7%; Pred. No. 0.17;  
Matches 139; Conservative 0; Mismatches 132; Indels 3; Gaps 1;  
QY 227 TCACAGGCGAAGAGACGCGGTAGTCCACCGGTGCGCGGTGAACGGTCCGGGGTGGCA 286  
Db 18890 TCTGGGCGGTAGAGACCGGTGCTCCAGGACGCCCGCGTGTGAGACCGCGGTGAGCG 18831  
QY 287 CCGGGATGTGCGTCCCGGGGTGAGCGCGCCCTGCTTTGGCCGGCGTGTGTCGCACACCC 346  
Db 18830 GGTGGCCGGGGTACCGGTGCGAGCAGGCGCGCGCGGTGCGCGTGC 18771  
QY 347 AGGAGACAGGTCCGGACCGATCCGAGAGTTCTGACGGTGGCGCTTTCCCGGTG 406  
Db 18770 AGCGGCCACGTCCG---CGTGGCGCCCAAGGCGCCAGTTCCGCGACAGTTCCGCG 18714  
QY 407 CGCGTACGCCACACCGCTTGCCTTCGCGCTTTCAGCGAGTTTCAGCAGGCGGAGCAGT 466  
Db 18713 CGCGGGGGCTGCGGGCGCTTCGGTGTGTCAGGACGAGCGCGCGGCGACCGTCCGCG 18654  
QY 467 CGTGGCGATCCCTCGACGTGCGCGGCGGACCT 500  
Db 18653 TGACAGATGCGGGCGCCCTCCGCGCGAGGCT 18620

## RESULT 39

US-08-576-626A-2/c  
; Sequence 2, Application US/08576626A  
; Patent No. 5998194  
; GENERAL INFORMATION:  
; APPLICANT: Summers, R.G.  
; APPLICANT: Katz, L.  
; APPLICANT: Donadio, S.  
; APPLICANT: Staver, M.J.  
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR  
; TITLE OF INVENTION: BIOSYNTHESIS GENES  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA: US/08/576,626A  
; FILING DATE: 21-DEC-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Diane Casuto  
; REGISTRATION NUMBER: P-40,943  
; REFERENCE/DOCKET NUMBER: 5857.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (847) 938-3137  
; TELEFAX: (847) 938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8051 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-576-626A-2

Query Match 9.3%; Score 46.6; DB 2; Length 8051;  
Best Local Similarity 49.8%; Pred. No. 0.18;

Matches 118; Conservative 0; Mismatches 119; Indels 0; Gaps 0;  
QY 1 GGTATCCCGACCGTGTCCCGGAACAACGAGTTCGAGATACGGGAGAGGAACACCCCGGT 60  
Db 7517 GATGCCCCACCGCGCCACGAAACGCTCTCTTCAGCGGGAGACACGAAAGAGGCCGCC 7458  
QY 61 AGTCGGGTAGACGCTGGCGGCGAAGGCTACGGCCCTTCAGCGTTCAGCGGGGGCGG 120  
Db 7457 GGTCTCTGTCTGGGTTCGGGGTGAACTCGAAACCGCCCGCCAGCTCACTACTC 7398  
QY 121 ACACCGCGGGTCACTCGTTCAGTGTACGCGGGGAGTACAGATCCACTGTCCGCCA 180  
Db 7397 GCAATCTCGTTCACTTCGCGGAGGAGCGGGTGTCTCCACCCCTCGCGCATG 7338  
QY 181 GCGCGCGGAACTCTCTCTTCGCCATGATCTGTCGGGTGTTCAGGCGAAG 237  
Db 7337 TCTCTTCGAGTGTGTTGAGCGGTTCGCGTCAACGACTGTCGAAAGCGTTCAGGAAG 7281

## RESULT 40

US-09-266-965-98/c  
; Sequence 98, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456U1  
; CURRENT APPLICATION NUMBER: US/09/266,965  
; CURRENT FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 08/624,447  
; EARLIER FILING DATE: 1996-08-19  
; EARLIER APPLICATION NUMBER: PCT/US94/11279  
; EARLIER FILING DATE: 1994-10-06  
; EARLIER APPLICATION NUMBER: US 08/133,963  
; EARLIER FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 98  
; LENGTH: 819  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-266-965-98

Query Match 9.3%; Score 46.4; DB 4; Length 819;  
Best Local Similarity 46.6%; Pred. No. 0.19;

Matches 149; Conservative 0; Mismatches 171; Indels 0; Gaps 0;  
QY 176 CGCCAGCCCGCGGAACTCTCTCTTCGCCATGATCTGTCGGCGTGTTCAGGCGA 235  
Db 334 CGCGGCGCGCGGACCTCGTTCGGGGCGGCTTCTCGACAGCGCGGAGAACCGCGCG 275  
QY 236 AGAGCAGCGGTAGTCCACCGGTTCGCGGTGAACCGTTCGGGGTGCACCGGGATGT 295  
Db 274 CGTGGGGAGTTTCGCGACCGCGCGTGTGTAAGCGTTCGCTCCCGCGCGCGGT 215  
QY 296 GCGTGC CGGGGTGAGCCCGCCCTCTCTTGGCGCGGTCTGTCGCACACCGAGAGACA 355  
Db 214 GTGCGGAGAGAACCGCGGCCCCACAGACGTCTGTCGAGGAAGACACCGCGGCG 155  
QY 356 GGTCCGACCGATGCGCAGAGTTCTGTCAGGTGGCGCTCTTCGCGTCCGCGGTACG 415  
Db 154 CGAGGCGCGCGCGCCCGCGCGGTTCGCGGGCGCGCGCGCGCGCGCGGTTCGCCGA 95  
QY 416 CCACCAACCGCTTGCCTTCGCGCTTCGAGCGATTTCAGCAGGCGCGAGGTCTGTCGGA 475  
Db 94 CCAAGCGCAGCGGAGAGGCGCGCTCATCTCGCCTGTTCAGCGCGATCAGTTGATCACCG 35  
QY 476 TGCCTTCGACGTCCGCGCG 495

Db 34 CGTCGCCGCCGTCGTTGTGG 15

RESULT 41

US-09-252-991A-10644/c  
; Sequence 10644, Application US/09252991A

Patent No. 6551795  
GENERAL INFORMATION:

INFORMATION: Rubenfield et al.  
APPLICANT: Marc J. NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: ABRUENOSA FOR DIAGNOSTICS AND THERAPEUTICS  
TITLE OF INVENTION:

FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18  
 ; CURRENT APPLICATION NUMBER: US 60/074,788  
 ; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIORITY FILING DATE: 1998-07-22  
 ; PRIORITY FILING DATE: 1998-07-22  
 ; NUMBER OF SEQ ID NOS: 33142

```

; SEQ ID NO 10644
;
; LENGTH: 1029

```

TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa

Query Match: 9.3%; Score 46.4; DB 4; Length 1029;

Best Local Similarity 45.1%; Pred. No. 0.2;  
Matches 173; Conservative 0; Mismatches 211; Indels 0; Gaps 0

QY 116 GSCGGACACCGGGGGTCAGCTCGTCA CGTGTACGGGGGACGTACAGGATCCACTGTC 175

Db  
559 GCGGATCGCGCGGTGGGAGGTAGATCTCTCGCGATACAGCGCAGAGCGCTGC 500

QY 176 GGCAGCCGGGGGAATCCTCTGTCTCTTCGCGCATGATCTCGTCGGCGTGSTTCAGGGCA 235

Db 499 CGGCATCGTAGCCACGTCACAGCGCGCTGTCGGTGTTCATCTCGACCTGGCTGCGGTCTT 440

QY 236 AGAGCAGCGGTAGTCCACCGTGGCGGTGAACGCGTCCGGGTGCGACCGGATGT 295

Db 439 CCAGCACAGTCCGCCCTCGCCGACGGCAGTGAGGTAGTCGGCGCGAAGCGTCG 380

RESULT 42

US-09-252-991A-10539/c  
; Sequence 10539, Application US/09252991A

Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

```

; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-02-18



; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 76  
; LENGTH: 53500  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-266-965-76

Query Match 9.3%; Score 46.4; DB 4; Length 53500;  
Best Local Similarity 46.6%; Pred. No. 0.21;  
Matches 149; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 176 CGCCAGCCGGCGGAACTCTGCTCTTCGCCATGATCTCGTCGGGCTGTTCCAGGCA 235  
DB 19824 CGCGGGCGCGCGAGACTCGTTCGGGGCGCTTCTCGACAGCGGCGAGAACCGCGCG 19883  
QY 236 AGAGCAGCGGTAGTCCACCGCTCGGGGTGAACCGCTCCGGGTGCGCACCGGGATG 295  
DB 19884 CGGTGGGAGTTCGCGCAGCGCGCGCTGGGTGAAGCGCTCCGTCGCGCGCGCGCT 19943  
QY 296 GCGTCCGGGGGTGAGCGCGCTGTTGCGCGGCTGCTGCGACACCCAGGAGCA 355  
DB 19944 GTCGGGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20003  
QY 356 GGTCCGACCGATCCCGCAGAACTGCTCACCGTGGCGCTCTTCGCGCGCGCGCTAG 415  
DB 20004 CGAGGGCG 20063  
QY 416 CACACCGCGCTTGCCTCGCGCTTGAGCGAGTTCAGCGGCGCGCAGCGTGGTCCGA 475  
DB 20064 CCACGCGCAGCGGAGCGCGCGCTCATCTCGCGCTGTCAGCGCGATGATCATCC 20123  
QY 476 TGCCCTCGAGCTCGCGCGCG 495  
DB 20124 CGTCGCGCGCGCTGTTGCG 20143

RESULT 47  
US-09-252-991A-122/c  
; Sequence 122, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 122  
; LENGTH: 822  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-122

Query Match 9.2%; Score 46.2; DB 4; Length 822;  
Best Local Similarity 46.2%; Pred. No. 0.21;  
Matches 153; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 138 CGTCACGTGTACGGGGGAGTACAGGATCCACTGTCGCCAGCGCGGCGGAACTCCTG 197  
DB 712 CGCTCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 653  
QY 198 CTCCTTCGCCATGATCTCGTCGGGCTGTTCCAGGCGAAGACAGCGCGGTAGTCCACCG 257  
DB 652 GGCAGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 593  
QY 258 GTCCGGCGTGAACCGCTCCGGGTGCGCACCGGATGTCGTCCCGGGGTGAGCCGCG 317

DB 592 GACGGTCTCCACCGGAAGGGTCTCTCCAGAGAGCGCGAGCGGTAGTGTGATGTC 533  
QY 318 CTGCTTGGCGCGCGCTGTCGCACACCCAGGAGACAGGTTCGGAGACCGGTATGCCGAGAA 377  
DB 532 GTAGCCATCGCGCGGACCTTCGCGCGCAACGCCCTCCAGTTCCTCGACGCTGCGCAGGT 473  
QY 378 GTTGTACAGGTGCGGCTCTTCGCGCTGCGCGGTAGCCACACCGCTTCCCTCGGC 437  
DB 472 GATCCGCGCGCGCGCGCTTCGCCATGCGGTGACCATACCCCGGATCGTCAAGGC 413  
QY 438 CTTAGCGAGTTCAGCAGGCGGAGCAGGTG 468  
DB 412 GATCAGCGGTTCGCTGTAGCCGTGGCG 382

## RESULT 48

US-09-252-991A-148  
; Sequence 148, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 148  
; LENGTH: 936  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-148

Query Match 9.2%; Score 46.2; DB 4; Length 936;  
Best Local Similarity 46.2%; Pred. No. 0.21;  
Matches 153; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 138 CGTCACGTGTACGGGGGAGTACAGGATCCACTGTCGCCAGCGCGGCGGAACTCCTG 197  
DB 73 CGCTCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATC 132  
QY 198 CTCCTTCGCCATGATCTCGTCGGGCTGTTCCAGGCGAAGACAGCGGTAGTCCACCG 257  
DB 133 GGCAGGCGCTTCGCCAGCGGTGCGCACCGGATGTCGTCCGCGCGCGCGCGCGCTC 192  
QY 258 GTCCGGCGTGAACCGCTCCGGGTGCGCACCGGATGTCGTCCGCGCGCGCGCGCGCG 317  
DB 193 GACGGTCTCCACCGGAGGGTCTCTCCAGAGCAGCGCGCGCGCGCGGTAGTGTGATGTC 252  
QY 318 CTGCTTGGCGCGCGCTGCTGTCGCACACCCAGGAGACAGGTTCGCGACCGCGAGAA 377  
DB 253 GTAGCCATCGCGCGGACCTTCGCGCGCAACCGCTCCAGTTCCTCGACGCTGCGCAGGT 312  
QY 378 GTTGTACAGGTGCGGCTCTTCGCGCTGCGCGGTAGCCACACCGCTTCCCTCGGC 437  
DB 313 GATCCGCGCGCGCGCGCTTCGCCATGCGGTGACCATACCCCGGATCGTCAAGGC 372  
QY 438 CTTAGCGAGTTCAGCAGGCGGAGCAGGTG 468  
DB 373 GATCAGCGGTTCGCTGTAGCCGTGGCG 403

## RESULT 49

US-09-252-991A-129/c  
; Sequence 129, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 129  
LENGTH: 975  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-129

Query Match  
Best Local Similarity 46.2%; Score 46.2; DB 4; Length 975;  
Matches 153; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 138 GTACAGTGTAGGGGGGAGTACAGATCCACTGTCCGACCCCGGGGGAACCTCTG 197  
Db 936 CGCTCGCGGTGGCGCGCGAGCGCGCGCGAGGTTCGCGCGGGCGCGGAGGATATC 877  
QY 198 CTCCTTCCCATGATCTCTCGCGGTGTTCAGCGGAAGACAGCGCTAGTCCACCGC 257  
Db 876 GGCAGAGGCTCGCCAGCGAGTGGGACCGGACCGGCTCCGCGTTCGACGCACTC 817  
QY 258 GTCCGCGGTGAACCGTCCGGGTGGCGACCGGAGTGTGCGCGGGGTGAGCCGCGC 317  
Db 816 GACGGTCTCCACCGGAGGGTCTTCAGAGAGCGCGGACGCGTAGTGTGATGTC 757  
QY 318 CTGCTTGGCGCGTGTGTCGACACCGAGAGACGAGTCCGAGCGGATCGCGAGAA 377  
Db 756 GTAGGCCATCGCGGACCTTCGCGGCAAGCTCCAGTTCTTCGACGCTGCGCAGGT 697  
QY 378 GTTCGTACAGTGTGCGCTTTCGCGTGTGCGCGGTGAGCCACCGCTTGCCTCGGC 437  
Db 696 GATCGCGCGCGCGCGCTTTCGCGTGTGCGCGTGTGAGCCATCACCGCGAGTGTGCAAGGC 637  
QY 438 CTTGAGCGAGTTCAGCGGCGGAGCAGGTG 468  
Db 636 GATCAGCGGTTGCGCTGTAGCCGTGGCG 606

RESULT 50  
US-08-254-357-1/c  
Sequence 1, Application US/08254357  
Patent No. 5610041  
GENERAL INFORMATION:  
APPLICANT: Christopher R. Somerville,  
APPLICANT: Christiane Nawrath,  
APPLICANT: Yves Poirier  
TITLE OF INVENTION: Processes For Producing  
TITLE OF INVENTION: Polyhydroxybutyrate and Related  
TITLE OF INVENTION: Polyhydroxyalkanoates in the  
TITLE OF INVENTION: Plastids of Higher Plants  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESS: Ian C. McLeod  
STREET: 2190 Commons Parkway  
CITY: Okemos  
STATE: Michigan  
COUNTRY: USA  
ZIP: 48864  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 5.25 inch, 360 kb  
MEDIUM TYPE: storage  
COMPUTER: Acer  
OPERATING SYSTEM: MS-DOS (version 3.3)  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/254.357  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/108,193 and 07/732,243  
FILING DATE: August 17, 1993 and July 19, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Ian C. McLeod  
REGISTRATION NUMBER: 20,931  
REFERENCE/DOCKET NUMBER: MSU 4.1-222  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (517) 347-4100  
TELEFAX: (517) 347-4103  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1431 Base Pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Double  
MOLECULE TYPE: Linear  
TOPOLOGY: Linear  
DESCRIPTION: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Alcaligenes eutrophus  
IMMEDIATE SOURCE:  
LIBRARY: Genomic  
US-08-254-357-1

Query Match 9.2%; Score 46.2; DB 1; Length 1431;  
Best Local Similarity 46.6%; Pred. No. 0.22;  
Matches 184; Conservative 0; Mismatches 208; Indels 3; Gaps 1;

QY 105 GTCAGCGCGCGGAGACACCGCGGCTCAGCTCCTGCTACAGTACCGCGGGGACGTACAG 164  
Db 660 GGCATCGCCCATGCGGAACCATCGCGGAGCCCGGACGACGTCGCGGGGCGGCTCAT 601  
QY 165 GATCCATCTGTCGCGAGCGCGGGAATCTCTGCTCTTTCGCGCATGATCTGTCGGCGT- 223  
Db 600 GTTTTCTGCGCGCGCGGACACCATCTCTGCGCTGCGCGCATGATGCGCTTGGCGGC 541  
QY 224 --GGTTCCAGGCGAAGAGCAGCGCGTAGTCCACCGCTCGGCGGTGAACGCTCCGGGT 281  
Db 540 CAGCATCAGGCTTTCAGGCGCGGCGGACACCTTGTGATGCTATGCGCGGACCAT 481  
QY 282 GCGCACCGGGATGTGCTGCGCGGGGTGAGCGCGCTGTTGGCGCGGCTGCTGCGCA 341  
Db 480 CGCGCGAGCGCGCTTTCATCGCGCTGCGCGCTTCTGCGCGGCTTCTGCGCGGCT 421  
QY 342 CACCCAGGAGACGCTCGGACCGATGCGCGAGGTTGTCACGCTGGCGCTCTTTCG 401  
Db 420 CAGCCTCGCGCATGATGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361  
QY 402 CGTCGCGGTACGCGCACCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 461  
Db 360 CGCGCGCTTTCGATGACCGCGGACCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 301  
QY 462 CAGTCTGCTGCGGATGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 496  
Db 300 AAACCTTCCGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGG 266

Search completed: June 27, 2004, 19:41:33  
Job time : 73.1177 secs













Human; diagnostic; drug screening; forensics; gene mapping; biodiversity assessment; Parkinson's disease; Alzheimer's disease; neurodegenerative diseases; anaemia; platelet disorder; wound; burns; ulcers; osteoporosis; autoimmune disease; cancer; molecular weight marker; food supplement; antiparkinsonian; nootropic; neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerary; antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic; gene therapy; chromosome 13q34; gene; ss.

Homo sapiens.

WO2003029271-A2.

10-APR-2003.

24-SEP-2002; 2002WO-US030474.

24-SEP-2001; 2001US-0324631P.

(HYSE-) HYSEQ INC.

Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T; Zhou P, Chosh M, Wang D, Ma Y, Asundi V, Wang Z, Wang G; Haley-Vicente D, Dmanac RT;

WPI; 2003-371981/35.

P-PSDB; ADC31291.

New polynucleotide and polypeptide useful for diagnosing, preventing or treating conditions such as neurodegenerative diseases, anemias, platelet disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or cancer.

Claim 1; SEQ ID NO 402; 1185bp; English.

The invention relates to 971 novel human cDNA sequences (ADC29919-ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The invention also relates to nucleic acid sequences over 93% identical with the novel human cDNAs. The invention additionally encompasses expression vectors and host cells comprising a nucleic acid of the invention; the recombinant production of a polypeptide of the invention; an antibody against a polypeptide of the invention; a method of detecting polynucleotides or polypeptides of the invention; and methods of identifying a compound which binds to a polypeptide of the invention. The invention further discloses methods of preventing, treating or ameliorating a medical condition; kits comprising polynucleotide probes and/or monoclonal antibodies for carrying out the methods of the invention; methods for the identification of compounds that modulate the expression or activity of the polynucleotide and/or polypeptide; and 767 contig sequences corresponding to the cDNA sequences of the invention (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628-ADC33394). The nucleic acids and polypeptides of the invention are useful in diagnostics, drug screening, forensics, gene mapping, in the identification of mutations responsible for genetic disorders or other traits, for assessing biodiversity, and in producing many other types of data and products dependent on DNA and amino acid sequences. They are also used for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, anaemia, platelet disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or cancer. The nucleic acids may also be used as hybridisation probes or primers, and in the recombinant production of a protein. The polypeptides are also useful in generating antibodies, as molecular weight markers, and as food supplements. The present sequence represents a specifically claimed human cDNA sequence of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 6941 BP; 1361 A; 2228 C; 2015 G; 1337 T; 0 U; 0 Other;

Query Match 10.2%; Score 53.2; DB 9; Length 6941;  
Best Local Similarity 46.0%; Pred. No. 0.49;  
Matches 216; Conservative 0; Mismatches 253; Indels 1; Gaps 1;

QY 46 ACTGCCACTCCGGCGGGGTGACACAGGTGCGACAGGTGGAGTTGAGCAGCCAGCTCATCG 105  
Db 1348 AGTTCCGGCGCGCAGTAGTAAGAGCCAAATGTCGGGGTCTGTCGGCCACGACCCCATCAGG 1407  
QY 106 CGCCTGCGCGCGCATCGGAAAACACCGGGCCAGATCAGTGCAGCAGCGGATGCGCG 165  
Db 1408 TCCCCGGCGCGCGCGCCACACACACCTGCTCAACCTCCCGCCAGCAGCGGCTGG 1467  
QY 166 CCTCGATCTCGACGGTCCGGCGCGCAGCTCGATCTGTCGCCCGGGTCCCCACACACAGGGAA 225  
Db 1468 TCGCGCGCTCGCGCACCGACAGCTGGCGCCACCCCGCGCGGCGCAAGTGCAGCTGT 1527  
QY 226 ACTGCTCGTGGGAGCGGAGCGCCCGCCAGCCCGGACAGCTCGTCAGGGTGGCGTCAACGA 285  
Db 1528 GCGGGTGGCAGCCCGCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1587  
QY 286 CCTCTCGCGGGTCCGGGAGCGGAAACGCGCAGCGGTTCAGATCCCTGTCAGTGCATCGGCT 345  
Db 1588 CGCAGAGGCTGGTGTGCGGTGGGAGC-CCCTGAGCCCGCGGCGGTCGGCGGCGCC 1646  
QY 346 CAGTCCCGGTGCTCCCGCTTGGCTGGGAGGATAGCGGTTACAGACAGCGGACCAACG 405  
Db 1647 CTGAGCCGCTCGCACACACCTGAGCGGCGGCTGCGGCGGCGGCGGAGCAAGTGGCGCTG 1706  
QY 406 CGGCGCGGGTCCGGGCGGCGGTTTCAGCCGATCCGTCGATGACCGGCTGCGGGTCCGG 465  
Db 1707 CTGCGCGGAGGGGCGGCTGCACACAGCGCTCATGTCATGCTGCGCGTGGCGACTCG 1766  
QY 466 GCGGTGCGGAGATCCGTACCGCGCGGACCGCTCGGCGGCGGCGGCGGCGG 515  
Db 1767 CGCGCGCGCGCCAGCCAGCGCGGCTCCCTGTCGTCAGCAGCGGCGGCGGCGG 1816

# RESULT 7

ACF34536  
ID ACF34536 standard; DNA; 4638 BP.

AC ACF34536;

DT 15-OCT-2003 (first entry)

DE Gene encoding angiogenesis protein ENO171.

XX Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;  
XX antipsoriatic; antiarteriosclerotic; cardiast; vasotropic; angiogenesis;  
KW gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;  
KW diabetic retinopathy; cardiovascular disease; atherosclerosis;  
KW ischemic limb disease; coronary artery disease; gene; ds.

OS Homo sapiens.

XX WO2003027285-A1.

XX 03-APR-2003.

XX 19-SEP-2002; 2002WO-AU001282.

XX 27-SEP-2001; 2001AU-00007973.

PR 27-SEP-2001; 2001AU-00007974.

PR 11-OCT-2001; 2001AU-00008210.

PR 29-OCT-2001; 2001AU-00008532.

PR 13-NOV-2001; 2001AU-00008838.

PR 28-AUG-2002; 2002AU-00951032.

XX (BION-) BIONOMICS LTD.

XX Gamble JR, Hahn CN, Vadas MA;

XX WPI; 2003-354655/33.

XX P-PSDB; ABR64261.

PT New angiogenic genes and polypeptides, useful for diagnosing,

PT prognosticating or treating an angiogenesis-related disorder, e.g.  
 PT cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or  
 PT cardiovascular diseases.  
 XX  
 PS Claim 2; SEQ ID NO 91; 90pp; English.  
 XX  
 CC The invention relates to the isolation of novel genes (ACF34446-ACF34559)  
 CC encoding proteins (ABR64180-ABR64281) involved in the process of  
 CC angiogenesis. The nucleic acid molecules are useful in identifying and/or  
 CC obtaining full-length human genes involved in an angiogenic process. The  
 CC nucleic acid molecule, polypeptides or complexes encoded, cells or  
 CC genetically modified non-human animals derived from these are useful for  
 CC the screening of candidate pharmaceutical compounds used in treating  
 CC angiogenesis-related disorders. They are also useful for diagnosing,  
 CC prognosticating or treating an angiogenesis-related disorder, which  
 CC involves uncontrolled or enhanced angiogenesis or is a disorder in which  
 CC a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,  
 CC diabetic retinopathy, psoriasis or cardiovascular diseases such as  
 CC atherosclerosis), or involves inappropriately arrested or decreased  
 CC angiogenesis or is a disorder in which an expanding vasculature is of  
 CC benefit (e.g. ischemic limb disease or coronary artery disease). The  
 CC modulator of expression or activity of the polypeptide encoded by the  
 CC nucleic acid sequence is useful for manufacturing a medicament for the  
 CC treatment of an angiogenesis-related disorder. This sequence corresponds  
 CC to the gene encoding one of the novel angiogenic protein  
 XX  
 SQ Sequence 4638 BP; 1352 A; 957 C; 1008 G; 1321 T; 0 U; 0 Other;

Query Match 10.0%; Score 52; DB 7; Length 4638;  
 Best Local Similarity 47.9%; Pred. No. 0.83;  
 Matches 181; Conservative 0; Mismatches 195; Indels 2; Gaps 1;  
 QY 6 AGCAGCAACACGCGTGGTTCGCCCGCGCGGCGCTGAGCGGCTCGACCGAGCGGCGGTG 65  
 DB 141 AGCCCGCGGCACAGCGCGCGGCGGCTGAGCGGCTCGACCGAGCGGCGGCGGCG 200  
 QY 66 ACCAGGTGACACAGTGGAGTTGAGCGAGCGAGCTATCGCGGCTGCGCGCGGATGCGG 125  
 DB 201 CCGCGCTCTCTCTGCTCGCGGCGGCGAGCGCGCGCGCGCGGCGGCGGCGGCGGCG 260  
 QY 126 AAACACCGCGGCGAGGATCAGCTGACGCGGCGATGCGCGCTCGATCTCGACGCTGCGG 185  
 DB 261 AGACCGGCTCTCGCGCGCGCGCGGCTCTGAGGAGCGGCGGCGGCGGCGGCGGAG 320  
 QY 186 CGAGCTGATCTGCTCGCGCGGCTTCCACACCGAGGGAACCTGCTCGGTGGGAGCGGCG 245  
 DB 321 AGCGCGCGCGCTGGGAGCGGCG--TCGGGCGCGCGCGGGAAGTGCTTGGCGGCGGCGG 378  
 QY 246 AGCCCGAGCGGCGACAGCTGCTGACGCTGCGGTGCGGTCAAGGAGCTCTCGCGGCTCGGAGCG 305  
 DB 379 GAGCGCGGCTGCGCGCGCGCGCGGCTGCGCGGAGCGAGCGGAGCGGAGCGGCGGCGG 438  
 QY 306 GAACCGCGCGGCTCAGATCCCTGTCTAGTCGCATCGGCTCAGTCCGCGGTGTCCTCCCTT 365  
 DB 439 GAGCGGAGCGCGGCTCCGAGCGCGCGCGCGGCTGCGCGGCGGCTGCGCGCGCGCGCTCCTCT 498  
 QY 366 GGCCTGGGAGATAGCGG 383  
 DB 499 CTCGCGCGCGGCGGCGG 516

## RESULT 8

AA53491  
 ID AA53491 standard; DNA; 114955 BP.  
 XX  
 AC AA53491;  
 XX  
 XX 05-JUL-1999 (first entry)  
 DT Human adenosine A1 receptor antisense oligonucleotide fragment.  
 DE  
 XX Antisense oligonucleotide; multiple target; antisense treatment;  
 XX impaired respiration; inflammation; lung disease;  
 KW  
 KW

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;  
 KW acute asthma; allergy; asthma; impeded respiration;  
 KW respiratory distress syndrome; pain; cystic fibrosis;  
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;  
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
 KW prostate cancer; ss.  
 XX Synthetic.  
 OS  
 XX WO9913886-A1.  
 XX  
 XX 25-MAR-1999.  
 XX  
 XX 17-SEP-1998; 98WO-US019419.  
 PF  
 XX 17-SEP-1997; 97US-0059160P.  
 PR  
 XX 09-JUN-1998; 98US-00093972.  
 PR  
 XX (UYEC-) UNIV EAST CAROLINA.  
 PA  
 XX Nyce JW;  
 FI  
 XX WPI; 1999-229400/19.  
 DR  
 XX  
 XX New antisense oligonucleotides used in treatment of, e.g. pulmonary  
 PT vasoconstriction.  
 PT  
 PS Disclosure; Page 37; 120pp; English.  
 XX  
 CC The specification describes antisense oligonucleotides (AA52869-X55271)  
 CC directed against at least 2 mRNAs selected from target genes, coding and  
 CC non-coding regions of RNAs corresponding to target genes, gene initiation  
 CC codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-  
 CC end and the juxta-section between coding and non-coding regions and all  
 CC segments of RNAs encoding proteins associated with one or more diseases,  
 CC conditions or mixtures. The antisense oligonucleotides may be derived  
 CC from sequences AA55272-74. These multiple target oligonucleotides  
 CC (specifically AA55180-271) can be used for the antisense treatment of  
 CC diseases and conditions. Typical diseases and conditions are those  
 CC associated with impaired respiration and inflammation, including lung  
 CC diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,  
 CC acute asthma, allergies, asthma, impeded respiration, respiratory  
 CC distress syndrome, pain, cystic fibrosis, pulmonary hypertension,  
 CC pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary  
 CC disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.  
 CC colon cancer, breast cancer, lung cancer, pancreatic cancer,  
 CC hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as  
 CC well as all types of cancers which may metastasize or have metastasized  
 CC to the lungs, including breast and prostate cancer  
 XX  
 SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

Query Match 9.9%; Score 51.6; DB 2; Length 114955;  
 Best Local Similarity 33.7%; Pred. No. 0.74;  
 Matches 168; Conservative 49; Mismatches 273; Indels 9; Gaps 1;  
 QY 26 GCCCGCGCGCGCGCATCACGAATGCCACTCCCGCGCGGCTGACAGGTCGACAGGTGGGA 85  
 DB 104171 GCSNNNDNNGTCGGCGCGGCGGCSNNNDNNTCGCGCGGCGGCSNNNCGCGGCSNNN 104230  
 QY 86 GTTGACGACCGAGCTCATCGCGCGCTCGCGCGCGGCGATCGGAAACACCGGCGGCGGATCAC 145  
 DB 104231 DNNCCGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCG 104290  
 QY 146 GTGACGACGCGGATGCGCGCGCTCGATC-----TCGACGCTCGGCGCGGCGGATCGAT 196  
 DB 104291 CGCGGCGCGGCGCSNNNDNNTCCGCGGCGCGGCGCGCGCGCGGCGGCGGCGGCGGCG 104350  
 QY 197 CTCGTGCGCGGCTCCACACCGAGGGGAACTGGCTCGGTGCGGAGCGGCGGCGGCGGCGG 256  
 DB 104351 CBGGCGCGGCGCGCGCGCGCGCGGCGGCGGCGGCSNNNDNNTCCGCGGCGGCGGCGGCGG 104410



PR 13-OCT-2000; 2000US-023924P.  
 PR 12-APR-2001; 2001US-0283296F.  
 PR 24-JUL-2001; 2001US-00910813.  
 XX (ECOP-) ECOPIA BIOSCIENCES INC.  
 PA Farnet CM, Zazopoulos E, Staffa A;  
 XX  
 DR WPI; 2002-435445/46.  
 DR P-PSDB; AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,  
 DR AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,  
 DR AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,  
 DR AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,  
 DR AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.  
 XX  
 PT Novel isolated ramoplanin biosynthetic pathway polypeptide useful for  
 PT chemically modifying biological molecule that is a substrate for a  
 PT polypeptide encoded by a ramoplanin biosynthesis gene cluster.  
 XX  
 XX Disclosure; Page 87-135; 212pp; English.  
 XX  
 CC The invention relates to an isolated ramoplanin biosynthetic pathway  
 CC polypeptide selected from a polypeptide of open reading frames (ORF) 1-  
 CC 32. The isolated polypeptides are useful for chemically modifying a  
 CC biological molecule that is a substrate for a polypeptide encoded by a  
 CC ramoplanin biosynthesis gene cluster, by contacting the biological  
 CC molecule with the isolated polypeptide, where the polypeptide chemically  
 CC modifies the biological molecule. The method comprises contacting the  
 CC biological molecule with at least two different polypeptides encoded by  
 CC ramoplanin ORFs 1-31. The polypeptides are useful for directing the  
 CC biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated  
 CC gene cluster comprising the ORFs is useful as a substrate for  
 CC bioengineering of antibiotic structures. An isolated polypeptide or its  
 CC encoding nucleic acid sequence is useful for generating derivatives of  
 CC ramoplanin, for improving production or for producing variants of other  
 CC antibiotics of the peptide class. The isolated polypeptides are useful  
 CC for synthesis of ramoplanin *in vivo* or *in vitro*, as an adenylation domain  
 CC in conjunction with other peptide synthetase modules and allowing the  
 CC incorporation of Thr into a peptide antibiotic precursor, for modifying  
 CC fatty acid structure and/or enhancing fatty acid incorporation into the  
 CC peptide antibiotic structure, for production of an hydroxyphenylglycine  
 CC (HPG)-containing peptide antibiotic, for enhancing secretion of  
 CC ramoplanin or its variants and derivatives, for enhancing uptake of  
 CC precursors for ramoplanin biosynthesis, for enhancing production of  
 CC ramoplanin products or its variants or derivatives, to chlorinate HPG of  
 CC a peptide antibiotic precursor, and for designing specific nucleotide  
 CC probes and primers for identifying and isolating putative lipopeptide  
 CC -producing microorganisms. This polynucleotide sequence represents the  
 CC 88421nt genomic DNA of a ramoplanin producing Actinoplanes sp.  
 CC microorganism of the invention  
 XX  
 SQ Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 U; 0 Other;  
 Query Match 9.8%; Score 51.2; DB 6; Length 88421;  
 Best Local Similarity 46.3%; Pred No. 0.89;  
 Matches 239; Conservative 0; Mismatches 273; Indels 4; Gaps 2;  
 7 GCACGACGACGCGGTGTCGCCCGCGCGCGCGGATGAGCGCGGTGCGCCCGCGAGCGCGA 67577  
 67518 GCACGACGACGTCGTGCGCGCGCGCGCGGATGAGCGCGGTGCGCCCGCGAGCGCGA  
 67 67518 GCACGACGACGTCGTGCGCGCGCGCGGATGAGCGCGGTGCGCCCGCGAGCGCGA 126  
 67578 ACTCGCCCGCGCGGTGCGCGCGCGCGGATGAGCGCGGTGCGCCCGCGAGCGCGA 67637  
 127 AACACCGCGCGCGGATGACGTCAGCAGCGCGGATGCGCGCTCGATCTCGACGTCGCGCC 186  
 67638 CGGTCGCGCGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 67697  
 187 GCAGTCGATCTGTCGTCGCGCGCGCTCCACACGAGCGGAACTGCTGCTGCTGCTGCTGCTGCT 246  
 67698 CGCCCTGCG 67757

QY 247 GCCCAGCCGCGGACAGCTGCTCCAGGCTGCGCTCAACGACCTCTCCGCGGTCCGCGAGCGG 306  
 DB 67758 GCCGCGCTCGCGCGCGCGCGCTGCGCGCTGCGCGAGCGCGCGCGCGCGCGCGCGCG 67817  
 QY 307 AAACGGCAGCGCTCAGATCCCTGTCAGTCGCTCAGTCGCTCAGTCGCTGCTGCTGCTG 366  
 DB 67818 CGAGCTGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 67875  
 QY 367 GCCTGGGAGGATAGCGGTTTCAACGACGAGCGGACACACGCGCGCGCGCGCGCGCGCG 424  
 DB 67876 GCCAAGGTCCTCCGCCGATACGAGCGCGCGCGCGCTGCGGGCTCGCGCGCGCGCG 67935  
 QY 425 TTACGCGGATCCGCTCGATGACACGAGCGCTGCGGGGTTCGCGCGGTTCGCGCGGATCCGTA 484  
 DB 67936 TTGGCGCGCGCGCGCGCTGCGGTGTCGCGCGCGCGCTCCAGCGCGCGCTCGCGCTCCG 67995  
 QY 485 CCCTCCGCGACCGCTCG 520  
 DB 67996 GCCGCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 68031  
 RESULT 10  
 ABZ66676  
 ID ABZ66676 standard; DNA; 975 BP.  
 XX  
 AC ABZ66676;  
 XX  
 DT 21-MAR-2003 (first entry)  
 XX  
 DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 14.  
 XX  
 KW Orthosomycin; biosynthesis; evernimicin; avilamycin; gene; ds.  
 XX  
 OS Streptomyces mobaraensis.  
 XX  
 FN WO200279505-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 28-MAR-2002; 2002WO-CA000432.  
 XX  
 PR 28-MAR-2001; 2001US-0279095P.  
 PR 30-MAR-2001; 2001US-0279709P.  
 PR 20-APR-2001; 2001US-0285214P.  
 XX  
 PA (ECOP-) ECOPIA BIOSCIENCES INC.  
 XX  
 PI Farnet CM, Zazopoulos E, Staffa A;  
 XX  
 DR WPI; 2003-058435/05.  
 DR P-PSDB; ABP99213.  
 XX  
 PT Identifying orthosomycin biosynthetic gene, gene fragment or gene  
 PT cluster, by detecting presence of nucleic acid sequence corresponding to  
 PT 17 of flammamycins protein families.  
 XX  
 PS Example 2; Page 173-174; 51pp; English.  
 XX  
 CC The invention relates to identifying orthosomycin biosynthetic genes and  
 CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
 CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
 CC ABP99362). The method is useful for identifying an orthosomycin  
 CC biosynthetic gene, gene fragment or gene cluster, especially an  
 CC evernimicin-type or avilamycin-type orthosomycin biosynthetic gene,  
 CC gene fragment or gene cluster. The method is useful for detecting the  
 CC presence of any organism that contains DNA for the production of  
 CC orthosomycins (both evernimicin-type orthosomycins and avilamycin-type  
 CC orthosomycins) regardless of the level at which genes for orthosomycin  
 CC production are expressed by the organism or the amount of orthosomycin  
 CC produced by the organism. This allows for the detection of new  
 CC orthosomycin natural products, not produced by the organism  
 XX  
 SQ Sequence 975 BP; 108 A; 378 C; 378 G; 111 T; 0 U; 0 Other;



Query Match 9.7%; Score 50.4; DB 7; Length 1209;  
Best Local Similarity 48.8%; Pred. No. 1.8; Mismatches 201; Indels 4; Gaps 2;  
Matches 195; Conservative 0;

QY 111 TCGCGCGCATCCGAAACACCGGCGCAGGATCAGTCAGCAGCGCGATCGCGCTCG 170  
DB 989 TCGCGGTGTGCGGAAACAGCGGGCGGCGGACCATGGGCACCCCGCGCAGCGCTCG 930  
QY 171 ATCTCGACGGTGGCGCGAGTCGATCTGTGCGCGGCTCCACACAGGGGAACCTGG 230  
DB 929 CGCAGCTGTTGGCCCGCGGTGGGTGACGAACAGCTGCGCGCATCCAGAGCAGCGCG 870  
QY 231 CTCGCTGGCAGGGCAGCGCCCGCGGAGACAGCTGTCAGGCTGCGCTCAGGACCTCT 290  
DB 869 TGGGGCGCCGCTCGACACAGCGCAGCTGCGCGCGCGCGCGGAGCG--CGGTGAC 813  
QY 291 CGCGGTTCGGAGCGGAAACAGCGGCTCAGTCCCTGTGATCGATCGGCTCAGTG 350  
DB 812 GGGATCCCCCGGTGACACACCGCGGCGAG-TCCACCTCGGACAGGCGCGCACGAG 754  
QY 351 CGGTGCTCCCTTGGCCTGGAGGATAGCGGTTCACGACGAGCGGACACACGCGCGGC 410  
DB 753 TCGCGCAGCGCTCCACCGGTTCGGGACGAGCGGCTCAGTCTGCGGCGCCACAG 694  
QY 411 GGGGGCGGGGGTTCAGCGGATCCGCTCGATGACGAGCGGCTGCGGGGTGCGGGCGGT 470  
DB 693 GCGCGCAGCGGATCCGACCGCGCCCGGTACACACGCGGCGGAGCGGTCCAGTCCGC 634  
QY 471 CGCGGAGATCCGTACCGCGCGGACCGCTTCGCGCCAGCGCC 510  
DB 633 CACCCACGGGGCAGCGACCGCGCGCTCGGGCGACGCGCGC 594

RESULT 13  
ABZ77318/c  
ID ABZ77318 standard; cDNA; 2684 BP.  
XX AC ABZ77318;  
XX DT 28-MAY-2003 (first entry)  
XX DE Nucleotide sequence of a murine ACDK3 polypeptide.  
XX KW Human; ACDK; activator of cyclin-dependent kinase; cell division;  
XX KW Chromosome 10; ACDK3; cyclin dependent kinase; expressed sequence tag;  
XX KW EST; A1606540; gene; ss.  
XX OS Mus musculus.  
XX PH Key Location/Qualifiers  
XX FT CDS 72..2213  
XX FT /\*tag= a  
XX FT /product= "ACDK3"  
XX FT /trans\_except= (pos:2159..2144,aa:Thr)  
XX PN WO2003008557-A2.  
XX PD 30-JAN-2003.  
XX PF 19-JUL-2002; 2002WO-US023147.  
XX PR 19-JUL-2001; 2001US-0306835P.  
XX PA (UIFL ) UNIV FLORIDA.  
XX PI She J, Wang C, Kumar PG;  
XX DR WPI; 2003-239327/23.  
XX DR P-PSDB; ABP97757.  
XX PT New proteins and genes that are activators of cyclin-dependent kinase  
XX PT (ACDK), useful for regulating cell division, or for screening ACDK  
XX PT modulators for regulating the progression of a cell through the cell

PT cycle.  
XX Claim 25; Page 71-72; 103pp; English.  
XX CC The present sequence encodes a murine ACDK (activator of cyclin-dependent  
CC kinases) 3 polypeptide. This sequence is corresponds to expressed  
CC sequence tag (EST) A1606540. Human ACDK1 is localised to chromosome 10q23  
CC -10q24. ACDK polypeptides are predominantly localised within the nuclei  
CC of cells able to undergo mitosis. In cells that do not undergo mitosis,  
CC ACDK proteins are found in the cytoplasm. The ACDK protein or  
CC polynucleotide is useful for regulating cell division, and for regulating  
CC cyclin dependent kinase. They are also useful for screening modulators of  
CC ACDK expression or activity. These are useful for modulating ACDK  
CC expression or activity in cells to regulate the progression of a cell  
CC through the cell cycle

QY Sequence 2684 BP; 471 A; 834 C; 906 G; 473 T; 0 U; 0 Other;  
Best Local Similarity 9.7%; Score 50.4; DB 7; Length 2684;  
Matches 129; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
QY 4 GCACGACGACGACCGGTGTGCGCGCGCGCCATCAGAACTGCACTCGCGCGGGG 63  
DB 780 GCCCGCGCAACACGACGCGGGGACGCGGGGTGGCCCGCGCCCTACAGCAGCA 721  
QY 64 TGACACAGTTCAGCAGGTGGAGTTGAGCAGCCAGCTCATCGCGGCTGCGCGGATGC 123  
DB 720 CGGCGAGTGGCGCTGCGCGAGGCTGGCCAGAAAGCAGCAGCGCGCGCGCGCG 661  
QY 124 CGAAACACCGGCGCAGGATCAGTGCAGCAGCGGATGCGCGCTCGATCTCGACGCTCG 183  
DB 660 CCCAGCGCGCGCGGGTTCAGGCGCGCGCGCGCGCGCTCGGCTCGAGCGGCTCT 601  
QY 184 GCGCGAGTGCATCTCGTCCCGCGCTCCACACCGAGGAGAACTGGCTCGGTGGCAGCG 243  
DB 600 CGCGCAGCAGCTGCACCTCGCGCGCGCGCGCGCGCGCGCTCAGTGCAGCGCGCG 541  
QY 244 GCAGCGCGCGCGGACAGC 263  
DB 540 CCACCGCGCGCGGCGCAGC 521

RESULT 14  
AAL61224  
ID AAL61224 standard; DNA; 82746 BP.  
XX AC AAL61224;  
XX DT 22-SEP-2003 (first entry)  
XX DE Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
XX KW Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX OS Actinosynnema pretiosum.  
XX PN WO2003045312-A2.  
XX PD 05-JUN-2003.  
XX PF 21-NOV-2002; 2002WO-US037547.  
XX PR 21-NOV-2001; 2001US-0332158P.  
XX PA (UNIW ) UNIV WASHINGTON.  
XX PI Floss HG, Yu T, Leistner E;  
XX DR WPI; 2003-493374/46.  
XX PT Novel maytansinoid produced by bacterial host cell transformed with  
XX PT expression vector comprising open reading frame from ansamitocin gene





QY 297 TCGGAGCGGAAACCGGACGGCTCAGATCCCTGTGATCGGCTCAGTCCGGTTC 356  
Db 2392 CAGAGGTTTCAGCGCGGGAACCTGCTGCTCCGCTGGTGCACAGGGTGACCGGGACCTG 2451  
QY 357 GTCGCCCTTGGCTCGGAGGATAGCGGTTTCACGACGAGCGGCACCA 402  
Db 2452 GCCGCACACGACGCGCGCGAGAGGGTCTACACACGCGGCACCA 2497

RESULT 16  
AAL61221/c  
ID AAL61221 standard; DNA; 1554 BP.  
AC AAL61221;  
XX  
DT 22-SEP-2003 (first entry)  
XX  
DE Actinosynnema pretiosum transcriptional activator gene #2.  
KW Maytansinoid; ansamitocin; antitumour; transcriptional activator; gene;  
KW ds.  
XX  
OS Actinosynnema pretiosum.  
XX  
PN WO2003045312-A2.  
XX  
PD 05-JUN-2003.  
XX  
PF 21-NOV-2002; 2002WO-US037547.  
XX  
PR 21-NOV-2001; 2001US-0332158P.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Floss HG, Yu T, Leistner E;  
XX  
DR WPI; 2003-493374/46.  
XX  
PT Novel maytansinoid produced by bacterial host cell transformed with  
PT expression vector comprising open reading frame from ansamitocin gene  
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX  
PS Disclosure; Page 103-104; 160pp; English.  
XX  
CC The invention relates to maytansinoid produced by bacterial host cell  
CC transformed with expression vector comprising open reading frame from  
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
CC useful as a potent antitumor agent. The present sequence is A. pretiosum  
CC ansamitocin gene cluster II transcriptional activator gene  
XX  
SQ Sequence 1554 BP; 157 A; 520 C; 704 G; 173 T; 0 U; 0 Other;

Query Match 9.5%; Score 49.6; DB 7; Length 1554;  
Best Local Similarity 47.3%; Pred. No. 2.5;  
Matches 229; Conservative 0; Mismatches 239; Indels 16; Gaps 2;

QY 37 CCATCAGCAACTGCCACTCCGCGGGGTGACAGGTGACAGGTGAGTTGACAGCC 96  
Db 553 CCGACAGCAGCTGCTCGCACAGAGGCGGCCACGTCCGCGGGTCCGAGTACAGCAGC 494  
QY 97 AGCTCATCCGCGCTCGCGCGCATGCCGAAACACCGGCCAGGATCAGTGCAGCAGG 156  
Db 493 CGGACAGCGCAGGACGCGCGCACCGCGCGGGTCCGCGCCAGCCGACCT 434  
QY 157 CGATGCGCGCTCGATCTCGAGTGTGCGCGGAGCTCGATCTGTCGCGCGGCTCCACA 216  
Db 433 CCGCGAGCACCGCTCCGCGCGGTCCGCGCACGTGTGCGCGTCCCGCGCCACCAAGCAGT 374  
QY 217 CCAGGGGAACTGGCTCGGTGTCAGCGGAGCGGCCAGCGGACAGCTGTCAGGCTGG 276  
Db 373 CCAGAGACCGCGTTTCGGCCCTCGGCGCTGTCTCGCCCTGTGTGCGCGC-----GG 323  
QY 277 CBTACGGACCTTCTCGCGGGTCTGGAGCGGAAACGCGCACCGCTCAGATCCCTGTGATC 336

Db 322 AGTCCACGTCCTCCCGCCCGCGCGGT-----CCCCGGCTTGGCCGACCCGGGTAGCG 270  
QY 337 GCATCGGCTCAGTGGCGGTGCTCCCTTGGCTTGGAGGATAGCGGTTTCACGACGAGCG 396  
Db 269 CCAGCCAGCAGTGTCTCCAGCACCGGACCTCTCTCGCGCGCGCGGTGCGGCCGAGC 210  
QY 397 GCACCAAGCGCGCGCGGGGGGGTTCAGCCGATCCGCTCGATGACACGCGGCTGC 456  
Db 209 CCGCGCAGCACCGCGTCCGAGTCCGCGGTGCTCCCGGCCACGCGCAGGTGCTCCACG 150  
QY 457 GGGGTGCGGGCGGTCCGCGAGATCCGTACCGCCCGGACCGCTCCGCGCAGCGCGCGG 516  
Db 149 GTGCGCTGTGCGCACCGGACACCTGCTCCCGCGCGGAAACGCGCCAGCAGCGGTGTG 90  
QY 517 ATCC 520  
Db 89 TGCC 86

RESULT 17  
AAL61225  
ID AAL61225 standard; DNA; 11905 BP.  
XX  
AC AAL61225;  
XX  
DT 22-SEP-2003 (first entry)  
XX  
DE Actinosynnema pretiosum ansamitocin biosynthetic gene cluster II.  
XX  
KW Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX  
OS Actinosynnema pretiosum.  
XX  
PN WO2003045312-A2.  
XX  
PD 05-JUN-2003.  
XX  
PF 21-NOV-2002; 2002WO-US037547.  
XX  
PR 21-NOV-2001; 2001US-0332158P.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Floss HG, Yu T, Leistner E;  
XX  
DR WPI; 2003-493374/46.  
XX  
PT Novel maytansinoid produced by bacterial host cell transformed with  
PT expression vector comprising open reading frame from ansamitocin gene  
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX  
PS Disclosure; Page 153-160; 160pp; English.  
XX  
CC The invention relates to maytansinoid produced by bacterial host cell  
CC transformed with expression vector comprising open reading frame from  
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
CC useful as a potent antitumor agent. The present sequence is A. pretiosum  
CC ansamitocin biosynthetic gene cluster II  
XX  
SQ Sequence 11905 BP; 1346 A; 4416 C; 4697 G; 1446 T; 0 U; 0 Other;

Query Match 9.5%; Score 49.6; DB 7; Length 11905;  
Best Local Similarity 47.3%; Pred. No. 2.1;  
Matches 229; Conservative 0; Mismatches 239; Indels 16; Gaps 2;

QY 37 CCATCAGCAACTGCCACTCCGCGGGGTGACAGGTGACAGGTGAGTTGAGCAGCC 96  
Db 11352 CCGACAGCAGTGTCTCGCACAGGAGCGCGGCCACGTCCGCGGTCCGAGTACAGCAGC 11411  
QY 97 AGCTCATCCGCGCTGCGCGCGCATGCCGAAACACCGGCCAGGATCAGTGCAGCAGC 156  
Db 11412 CGGACAGCGCAGGACGCGCGGCCACCGCGCGCGGGTCCCGCGCCAGCGCAGCT 11471

157 CGATCGCGCTCGATCTCGACGCTGGCGGCGAGCTCGATCTCGTCCGCGGTCCACCA 216  
 11472 CCAGCAGCAGCTCGCTCGCGCGCTGGCGCAGCTGCTCGCGTCCCGGCGCAGCAGCT 11531  
 217 CCAGGGGAACTGCTCGCTGGCAGCGGCGAGCGGCGGCGGCGGCGGCGGCGGCGG 276  
 11532 CCAGCAGCAGCTCGCTCGCGCGCTGGCGGCTCGCTCGCGCTGGCGGCGGCGGCGG 11582  
 277 CFTACCGGACTCTCGCGGCTGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 336  
 11583 AGTCCAGCTCGCGCGCGCGCGCT-----CCCGGCTGGCGCAGCAGCGGCTAGGCG 11635  
 337 GCATCGCTCAGTCCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 396  
 11636 CCAGCAGCAGCTCGCTCGCAGCAGCGGCACTCTCGCGCGCGCGGCTCGCGCGCG 11695  
 397 GCACCAAGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 456  
 11696 CCGCGAGCAGCGCGCTCGGAGTGGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGG 11755  
 457 GGGTCCGGGCGGCTCGCGGAGATCCGTACCGCGCGGAGCGGCGGCGGCGGCGGCGG 516  
 11756 GTCCGCTGTCCGCAACCGCACTCGCGCGGCGGAGCGGCGGCGGCGGCGGCGG 11815  
 517 ATCC 520  
 11816 TGCC 11819  
 RESULT 18  
 ACA37544/c  
 ID ACA37544 standard; DNA; 1575 BP.  
 AC ACA37544;  
 XX  
 19-JUN-2003 (first entry)  
 XX  
 Prokaryotic essential gene #19201.  
 XX  
 Antisense; ds; prokaryotic essential gene; cell proliferation;  
 XX  
 drug design; gene.  
 XX  
 Mycobacterium avium.  
 OS  
 WO200277183-A2.  
 XX  
 03-OCT-2002.  
 XX  
 21-MAR-2002; 2002WO-US009107.  
 XX  
 21-MAR-2001; 2001US-00815242.  
 XX  
 06-SEP-2001; 2001US-00948993.  
 XX  
 25-OCT-2001; 2001US-0342923P.  
 XX  
 08-FEB-2002; 2002US-00072851.  
 XX  
 06-MAR-2002; 2002US-0362699P.  
 XX  
 (ELIT-) ELITRA PHARM INC.  
 XX  
 Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 WPI; 2003-029926/02.  
 DR  
 P-PSDB; ABU33674.  
 DR  
 New antisense nucleic acids, useful for identifying proteins or screening  
 XX  
 for homologous nucleic acids required for cellular proliferation to  
 PT  
 isolate candidate molecules for rational drug discovery programs.  
 XX  
 Claim 14; SEQ ID NO 25414; 1766pp; English.  
 PS  
 The invention relates to an isolated nucleic acid comprising any one of  
 XX  
 the 6213 antisense sequences given in the specification where expression  
 CC

of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1575 BP; 224 A; 523 G; 592 C; 236 T; 0 U; 0 Other;

Query Match 9.5%; Score 49.2; DB 7; Length 1575;  
 Best Local Similarity 50.0%; Fred. No. 2.9; Mismatches 0; Gaps 0;  
 Matches 123; Conservative 0; Indels 123; Indels 0; Gaps 0;  
 QY 44 GAACTGCCACTCCGCGGGGTGACACAGTCCAGGCTGACGAGTGGAGTTGAGCAGCGCTCAT 103  
 DB 1542 GATGTCCAGCAGCAGCGGTTGACCTCGGGGACCTCGTTGCTGATGCGGGTCCGATCCG 1483  
 QY 104 CCGCCCTCGCGCGCATGCCGAACACCGCGGCGGAGTACGTCGACGCGGATCGG 163  
 DB 1482 CTCAGACACCTCGTAGGGGACCGGGTCCAGTCCGCGGTCTGCGTCTCTCCTGCTGTCAC 1423  
 QY 164 CGCTTCGATCTCGACGCTCGCGCGGCGAGCTCGATCTCGTCCGCGCGCTCCACACCGAGGG 223  
 DB 1422 CGCGCGCAGCAGATCGGATGCCGTAGTGGGTGTCGCCCTGACCCCGCAGCGG 1363  
 QY 224 AAATCGCTCGGTGCGAGCGGCGAGCGCGGCGGAGCAGCTCGTCCAGGCTGGGCTCAG 283  
 DB 1362 GACCTCAGCAGCAGCAGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1303  
 QY 284 GACCTC 289  
 DB 1302 GAGTTC 1297

RESULT 19  
 ADA71938  
 ID ADA71938 standard; DNA; 2000 BP.  
 XX  
 AC ADA71938;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Rice gene, SEQ ID 5263.  
 XX  
 KW Plant; bacterial infection; fungal infection; viral infection; rice;  
 KW gene; ds.  
 XX  
 OS Oryza sativa.  
 XX  
 PN WO2003000898-A1.





FT		/tag= f	
FT		/product= "ORF6 protein"	
FT		/note= "No start codon"	
FT		/partial	
FT	CDS	41772..47633	
FT		/tag= g	
FT		/product= "ORF7 protein"	
FT		/note= "No start codon"	
FT		/partial	
FT	CDS	47635..49890	
FT		/tag= h	
FT		/product= "ORF8 protein"	
FT	CDS	49922..50938	
FT		/tag= i	
FT		/product= "ORF9 protein"	
FT	CDS	51234..52079	
FT		/tag= j	
FT		/product= "ORF10 protein"	
XX			
PN	WO200288176-A2.		
XX			
XX	07-NOV-2002.		
XX			
PF	26-APR-2002; 2002WO-CA000591.		
XX			
PR	26-APR-2001; 2001US-0286346P.		
XX			
PA	(ECOP-) ECOPIA BIOSCIENCES INC.		
XX			
PI	Farnet CM, Zazopoulos E, Staffa A, Yang X;		
XX			
XX	WPI; 2003-201222/19.		
DR	P-FSDB; AAE35484, AAE35485, AAE35486, AAE35487, AAE35488, AAE35489,		
DR	AAE35490, AAE35491, AAE35492, AAE35493.		
XX			
PT	Novel isolated or purified polypeptide involved in biosynthesis of		
PT	polyketide dorrigin or polyketide lactimidomycin, useful for preparing		
PT	dorrigin or lactimidomycin.		
XX			
PS	Claim 1; Page 85-113; 312pp; English.		
XX			
CC	The invention relates to novel proteins involved in the biosynthesis of		
CC	polyketide dorrigin (DORR) or lactimidomycin (LACT) biosynthesis by		
CC	microorganisms. Sequences of the invention allow direct manipulation of		
CC	dorrigin, lactimidomycin and related chemical structures via chemical		
CC	engineering of the enzymes involved in the biosynthesis of dorrigin and		
CC	lactimidomycin. They are useful for introducing chemical handles into		
CC	normally inert positions that permit subsequent chemical modifications		
CC	and facilitate the development of polyketides. The genes and proteins of		
CC	the invention can also be used to generate a focused library of analogues		
CC	around a polyketide lead candidate to fine-tune the compound for optimal		
CC	properties. They are useful for generating antibodies specific for the		
CC	polyketide biosynthesis. The present sequence is Streptomyces platensis		
CC	subspecies rosaceus DORR DNA		
XX			
SQ	Sequence 52101 BP; 7364 A; 20113 C; 17894 G; 6730 T; 0 U; 0 Other;		
	Query Match 9.4%; Score 49; DB 7; Length 52101;		
	Best Local Similarity 49.8%; Pred. No. 2.3;		
	Matches 124; Conservative 0; Mismatches 125; Indels 0; Gaps 0;		
QY	71 GTCAGCAGGTGGAGTTAGACGCCAGCTCAFCGCCGCTGCGCGGATGCCGAAACA 130		
Db	30119 GTGGCGCTGTGGGAGTTGGGATGAAGTCCACGAGGAGATCAGCAGGACGAAGTC 30060		
QY	131 CCGGGCCAGATCACGTGCAGCAGCGGATGCGCGCTCGATCTCGACGGTCGGCGCAG 190		
Db	30059 CAGCCGTGCCCGGGAAGACGTGCGGCAAGTGGCGGACGGTGCACCTTGGCCGCGAG 30000		
QY	191 CTCGATCTCGTCGCCGCTCCACACAGGGGAAATCGCTCGGTGGCAGCGCAGCCC 250		
Db	29999 GGTGGCCTCCAGTTGGGCTCGGTCTATCCGGGCGAGCGGGTGTGACTGAACACCATCG 29940		

QY	251 CAGCGGACAGCTCGTCAGGCTCGCTCAGGACCTCTCGGGGTGCGGAGCGAAAC 310		
Db	29939 CAGGTGACACCGCTCGATGCGCGGTAGCGCGGACGACCTCGTCCGCTGCGGGC 29880		
QY	311 GCGCAGGC 319		
Db	29879 GAGCGGTC 29871		
	RESULT 23		
	AAS08693/c		
ID	AAS08693 standard; DNA; 109519 BP.		
XX			
AC	AAS08693;		
XX			
DT	11-SEP-2003 (revised)		
DT	26-SEP-2001 (first entry)		
XX			
DE	Micromonospora DNA encoding biosynthetic enzymes for Evernimycin.		
XX			
KW	Evernimycin; antibiotic; bottle-neck gene; orthomycin; fermentation;		
KW	ds.		
XX			
OS	Micromonospora sp. ATCC 39149.		
XX			
FH	Key Location/Qualifiers		
FT	CDS complement(132..1382)		
FT	/tag= a		
FT	/product= "EvdA"		
FT	complement(1389..1394)		
FT	/tag= b		
FT	complement(1490..2611)		
FT	/tag= c		
FT	/product= "EvdB"		
FT	complement(2618..2622)		
FT	/tag= d		
FT	complement(2622..3860)		
FT	/tag= e		
FT	/product= "EvdC"		
FT	complement(3867..3870)		
FT	/tag= f		
FT	4134..4138		
FT	/tag= h		
FT	4143..5312		
FT	/tag= g		
FT	/product= "EvdD"		
FT	5309..6235		
FT	/tag= i		
FT	/product= "EvdE"		
FT	6226..6229		
FT	/tag= k		
FT	6232..7275		
FT	/tag= j		
FT	/product= "EvdF"		
FT	7272..8327		
FT	/tag= l		
FT	/product= "EvdG"		
FT	8333..8336		
FT	/tag= n		
FT	8342..9364		
FT	/tag= m		
FT	/product= "EvdH"		
FT	complement(9463..10224)		
FT	/tag= o		
FT	/product= "EvdI"		
FT	complement(10232..10235)		
FT	/tag= p		
FT	10424..11176		
FT	/tag= q		
FT	/product= "EvdJ"		
FT	12027..12455		
FT	/tag= r		
FT	/product= "EvdK"		
FT			

FT	/product= "EvrQ"
FT	complement(38072..38566)
FT	/*tag= at
FT	/*tag= au
FT	/*tag= av
FT	/*tag= ax
FT	/*tag= ay
FT	/*tag= az
FT	/*tag= ba
FT	/*tag= bb
FT	/*tag= bd
FT	/*tag= be
FT	/*tag= bg
FT	/*tag= bf
FT	/*tag= bh
FT	/*tag= bi
FT	/*tag= bj
FT	/*tag= bk
FT	/*tag= bl
FT	/*tag= bm
FT	/*tag= bn
FT	/*tag= bo
FT	/*tag= bp
FT	/*tag= bq
FT	/*tag= br
FT	/*tag= bs
FT	/*tag= bt
FT	/*tag= bu
FT	/*tag= bv
FT	/*tag= bw
FT	/*tag= bx
FT	/*tag= by
FT	/*tag= bz
FT	/*tag= ca
FT	/*tag= cb
FT	/*tag= cc
FT	/*tag= cd
FT	/*tag= ce
FT	/*tag= cf
FT	/*tag= cg
FT	/*tag= ch
FT	/*tag= ci
FT	/*tag= cj
FT	/*tag= ck
FT	/*tag= cl
FT	/*tag= cm
FT	/*tag= cn
FT	/*tag= co
FT	/*tag= cp
FT	/*tag= cq
FT	/*tag= cr
FT	/*tag= cs
FT	/*tag= ct
FT	/*tag= cu
FT	/*tag= cv
FT	/*tag= cw
FT	/*tag= cx
FT	/*tag= cy
FT	/*tag= cz
FT	/*tag= da
FT	/*tag= db
FT	/*tag= dc
FT	/*tag= dd
FT	/*tag= de
FT	/*tag= df
FT	/*tag= dg
FT	/*tag= dh
FT	/*tag= di
FT	/*tag= dj
FT	/*tag= dk
FT	/*tag= dl
FT	/*tag= dm
FT	/*tag= dn
FT	/*tag= do
FT	/*tag= dp
FT	/*tag= dq
FT	/*tag= dr
FT	/*tag= ds
FT	/*tag= dt
FT	/*tag= du
FT	/*tag= dv
FT	/*tag= dw
FT	/*tag= dx
FT	/*tag= dy
FT	/*tag= dz
FT	/*tag= ea
FT	/*tag= eb
FT	/*tag= ec
FT	/*tag= ed
FT	/*tag= ee
FT	/*tag= ef
FT	/*tag= eg
FT	/*tag= eh
FT	/*tag= ei
FT	/*tag= ej
FT	/*tag= ek
FT	/*tag= el
FT	/*tag= em
FT	/*tag= en
FT	/*tag= eo
FT	/*tag= ep
FT	/*tag= eq
FT	/*tag= er
FT	/*tag= es
FT	/*tag= et
FT	/*tag= eu
FT	/*tag= ev
FT	/*tag= ew
FT	/*tag= ex
FT	/*tag= ey
FT	/*tag= ez
FT	/*tag= fa
FT	/*tag= fb
FT	/*tag= fc
FT	/*tag= fd
FT	/*tag= fe
FT	/*tag= ff
FT	/*tag= fg
FT	/*tag= fh
FT	/*tag= fi
FT	/*tag= fj
FT	/*tag= fk
FT	/*tag= fl
FT	/*tag= fm
FT	/*tag= fn
FT	/*tag= fo
FT	/*tag= fp
FT	/*tag= fq
FT	/*tag= fr
FT	/*tag= fs
FT	/*tag= ft
FT	/*tag= fu
FT	/*tag= fv
FT	/*tag= fw
FT	/*tag= fx
FT	/*tag= fy
FT	/*tag= fz
FT	/*tag= ga
FT	/*tag= gb
FT	/*tag= gc
FT	/*tag= gd
FT	/*tag= ge
FT	/*tag= gf
FT	/*tag= gg
FT	/*tag= gh
FT	/*tag= gi
FT	/*tag= gj
FT	/*tag= gk
FT	/*tag= gl
FT	/*tag= gm
FT	/*tag= gn
FT	/*tag= go
FT	/*tag= gp
FT	/*tag= gq
FT	/*tag= gr
FT	/*tag= gs
FT	/*tag= gt
FT	/*tag= gu
FT	/*tag= gv
FT	/*tag= gw
FT	/*tag= gx
FT	/*tag= gy
FT	/*tag= gz
FT	/*tag= ha
FT	/*tag= hb
FT	/*tag= hc
FT	/*tag= hd
FT	/*tag= he
FT	/*tag= hf
FT	/*tag= hg
FT	/*tag= hh
FT	/*tag= hi
FT	/*tag= hj
FT	/*tag= hk
FT	/*tag= hl
FT	/*tag= hm
FT	/*tag= hn
FT	/*tag= ho
FT	/*tag= hp
FT	/*tag= hq
FT	/*tag= hr
FT	/*tag= hs
FT	/*tag= ht
FT	/*tag= hu
FT	/*tag= hv
FT	/*tag= hw
FT	/*tag= hx
FT	/*tag= hy
FT	/*tag= hz
FT	/*tag= ia
FT	/*tag= ib
FT	/*tag= ic
FT	/*tag= id
FT	/*tag= ie
FT	/*tag= if
FT	/*tag= ig
FT	/*tag= ih
FT	/*tag= ii
FT	/*tag= ij
FT	/*tag= ik
FT	/*tag= il
FT	/*tag= im
FT	/*tag= in
FT	/*tag= io

QY	29	82792	82733
Db	CCGCGCCGCCATACAGAACTGCCACTCCGCGCGGGTGACAGGTCGACCGAGTGGGAGTT	CCCTTCGAGCTAGACGGTCGGCGGCTCGTCCGGCTGCCGAGCGCCGAGAGCGGGAACTC	

```
QY      89 GAGCAGCCAGCTCATCGCGCGCTGCGCGGCGATGCCGAAACACCGCGGCCAGGATCACGTG 148
      |||||
Db      82732 GAGCAGCAGACGTCGCGGCTTCGCGCGCGCAGCGCGCGCGCGCGCGCAGCAGCAGC 82673
      |||||
QY      149 CAGCAGCGCGATCGCGCGCTCGATCTCGACGCTCGCGCGCGCGAGTCGATCTCGTCCCGG 208
      |||||
Db      82672 CGGTACGCTGATGTTGTCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAT 82613
      |||||
QY      209 CTCGCCACACCGAGGAACTGCTCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 253
      |||||
Db      82612 CACCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGAG 82568
      |||||

RESULT 24
ID      AAD54645/c
XX      AAD54645 standard; DNA; 113193 BP.
AC      AAD54645;
XX      26-JUN-2003 (first entry)
XX      Streptomycetes nodosus amphotericin (amph) biosynthetic gene cluster.
DE      Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; gene; ds.
KW      Streptomycetes nodosus.
XX      FH Key Location/Qualifiers
      FT CDS complement(4. .1824)
      FT FT /*tag= a
      FT FT /product= "ABC transporter encoded by S. nodosus amphg
      FT FT gene"
      FT FT complement(1805. .3628)
      FT FT /*tag= b
      FT FT /product= "ABC transporter encoded by S. nodosus amphH
      FT FT gene"
      FT FT 3840. .4874
      FT FT /*tag= c
      FT FT /product= "GDP-mannose dehydratase encoded by S. nodosus
      FT FT amphDII gene"
      FT FT 5042. .33574
      FT FT /*tag= d
      FT FT /product= "Polyketide synthase multienzyme housing
      FT FT extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
      FT FT nodosus amphI gene"
      FT FT 33584. .50518
      FT FT /*tag= e
      FT FT /product= "Polyketide synthase multienzyme housing
      FT FT extension modules 15, 16 and 17 encoded by S. nodosus
      FT FT amphJ gene"
      FT FT 50571. .56675
      FT FT /*tag= f
      FT FT /product= "Polyketide synthase multienzyme housing
      FT FT extension modules 18 and thioesterase encoded by S.
      FT FT nodosus amphK gene"
      FT FT 56829. .58019
      FT FT /*tag= g
      FT FT /product= "Cytochrome P450 encoded by S. nodosus amphL
      FT FT gene"
      FT FT 58139. .58648
      FT FT /*tag= h
      FT FT /product= "ORF1, hypothetical protein"
      FT FT complement(58756. .59610)
      FT FT /*tag= i
      FT FT /product= "ORF2, hypothetical protein"
      FT FT 59869. .61470
      FT FT /*tag= j
      FT FT /product= "ORF3, hypothetical protein"
      FT FT /transl_except= (pos:59869. .59871, aa:Met)
      FT FT complement(61798. .61995)
      FT FT /*tag= k
      FT FT /product= "Ferredoxin encoded by S. nodosus amphM gene"
```

```
FT CDS complement(62051. .63250)
FT FT /*tag= l
FT FT /product= "Cytochrome P450 encoded by S. nodosus amphN
FT FT gene"
FT CDS complement(63250. .64308)
FT FT /*tag= m
FT FT /product= "NDP-sugar aminotransferase encoded by S.
FT FT nodosus amphDII gene"
FT CDS complement(64324. .65775)
FT FT /*tag= n
FT FT /product= "Glycosyl transferase encoded by S. nodosus
FT FT amphDI gene"
FT FT /transl_except= (pos:65773. .65775, aa:Met)
FT CDS 60801. .70319
FT FT /*tag= o
FT FT /product= "Polyketide synthase multienzyme housing
FT FT loading module encoded by S. nodosus amphA gene"
FT CDS 70366. .79938
FT FT /*tag= p
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 1 and 2 encoded by S. nodosus by amphB
FT FT gene"
FT CDS 79956. .112709
FT FT /*tag= q
FT FT /product= "Polyketide synthase multienzyme housing
FT FT extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
FT FT nodosus by amphC gene"
XX
XX WO200297082-A2.
XX
XX 05-DEC-2002.
XX
XX 27-MAY-2002; 2002WO-1E000071.
XX
XX 31-MAY-2001; 2001IE-00000527.
XX
XX (UYDU-) UNIV COLLEGE DUBLIN.
XX
XX Caffrey JP;
XX
XX WPI; 2003-201271/19.
XX
XX P-PSDS; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
XX AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
XX AAE36129, AAE36130, AAE36131, AAE36132.
XX
XX Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
XX for preparing amphotericin derivative or analog antibiotic agent with
XX altered properties, in biosynthesis of polyketide other than
XX amphotericin.
XX
XX Claim 1; Page 52-114; 276pp; English.
XX
XX The invention relates to the gene cluster encoding the polypeptides
XX responsible for the biosynthesis of the polyene antibiotic amphotericin
XX (amph) of Streptomyces nodosus. Polynucleotides of the invention are
XX useful for preparing amphotericin derivatives or analogue antibiotic
XX agents with altered properties and in the biosynthesis of polyketides
XX other than amphotericin. amphDII, amphDII or amphDI mutants are useful
XX for producing amphotericin derivatives glycosylated with alternative
XX sugars; amphDII or amphDII gene sequences are useful in engineered
XX biosynthesis of perosaminyl-amphoteronolide B; amphDII or amphDII and
XX perosaminyl-16-descarboxyl-16-methyl amphoteronolide B; amphDII, amphDII
XX and amphDI gene sequences are useful for preparing polypeptides capable
XX of addition of mycosamine to a polyketide other than amphoteronolide A or
XX B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
XX The present sequence is S. nodosus amph biosynthetic gene cluster
XX
XX Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;
XX
XX Query Match 9.4%; Score 49; DB 7; Length 113193;
XX Best Local Similarity 48.1%; Pred. NO. 2.2;
XX Matches 139; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
```



QY 40 TCACGAAGTCCACTCCGCGGGTGAACAGGTGACAGGTCAGAGTGGAGTTGACAGCCAGC 99  
| | | | |  
Db 21770 TCACCCGACAGCAGTCTGTCGCGGCCCTCTGTCGCGAAGTTCGCGGAGGCGGAGCA 21711  
| | | | |  
QY 100 TCATCGCGGCTGCGCGGATGTCGGAACACCGGGCCAGGATCAGTGCAGCAGCGCGA 159  
| | | | |  
Db 21710 CGGGCTCTCTCGCGCGGACGACGACGAGCGCGGCGCCGTTGACGGCGGCCACGGAGA 21651  
| | | | |  
QY 160 TGGCGGCTCGATTCGACCGTGGCGCGAGTCTGATCTGTCGCGCGGCTCCACACCA 219  
| | | | |  
Db 21650 CTTGTCGGGAGGCGCTCCAGGTGCGGCGACCTGCTCTCGGGGCGCTCCAGGCGCA 21591  
| | | | |  
QY 220 GGGGAACTGCTCGGTGGCAGCGGACGCGCCAGCGCGGACAGTCTGTCAGGCTGGCGT 279  
| | | | |  
Db 21590 CCATCGCGCGCTCTCGGCGAGTCTGTCATGACGCGCGCGCGGCCACAGGCTGC 21531  
| | | | |  
QY 280 CACGACCTCTCGCGGTCGCGAGCGGAAACGCGACGCGTTCAGATCC 328  
| | | | |  
Db 21530 ACGGTCGTCGAGTCCGAGGACGCGCGGACATGCGCGGCGGATCTC 21482  
| | | | |

RESULT 25

ACA26655 standard; DNA; 2208 BP.

XX AC

XX AC

XX AC

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #9312.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Burkholderia mallei.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US09107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX P-PSDB; ABU22785.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 14525; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

XX Sequence 2208 BP; 398 A; 737 C; 767 G; 306 T; 0 U; 0 Other;

XX Query Match 9.4%; Score 48.8; DB 7; Length 2208;

XX Best Local Similarity 49.6%; Pred. No. 3.3;

XX Matches 125; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

XX QY 65 GACGAGTTCGACAGTGGGAGTTGAGCAGCCAGTATCGCCGCTCGCGCGCATGCC 124

XX Db 826 GACTGGGGGACACGTGAGCACTTCTGCGTGGCGGACGCTGCTCCGAAATCCG 885

XX QY 125 GAAACACCGCGCCAGGATCAGTGCAGCAGCAGCGATGCGCCCTCGATCTCGACGGTGG 184

XX Db 886 AACGGCAGCATCTCGGACGACCTGTACACGGCCGATCGAACTTCGACCATTCGCGCAAG 945

XX QY 185 CGCAGCTCGATCTCTGTCGCGCGCTCCACACGAGGGAACCTGCTCGTGGCAGCGG 244

XX Db 946 AAGCAGTGGTCTGCGGTATCAGTTGAGCACAAGCTGAACCGGTGTGACGCTGCGG 1005

XX QY 245 CAGCCCCAGCGGGACAGCTGCTCCAGGCTGGCTCGCGACCTCTCGCGGCTCGGGAGC 304

XX Db 1006 CAGAACGTCGCTGGATGATGACCTGCTGCTGACGACGCGTCCGCTAGCGGGCGGCTC 1065

XX QY 305 GGAACCGCGCAC 316

XX Db 1066 GACGACCGCAGAC 1077

RESULT 26

AAL61219/c

ID AAL61219 standard; DNA; 795 BP.

XX AC

XX AAL61219;

XX DT 22-SEP-2003 (first entry)

XX DE Actinosynnema pretiosum Asm46 gene.

XX XW Maytansinoid; ansamitocin; antitumour; Asm46; gene; ds.

XX OS Actinosynnema pretiosum.

XX PN WO2003045312-A2.

XX PD 05-JUN-2003.

XX PF 21-NOV-2002; 2002WO-US037547.

XX PR 21-NOV-2001; 2001US-0332158P.

XX PA (UNIW ) UNIV WASHINGTON.

```
PI Floss HG, Yu T, Leistner E;
XX WPI; 2003-493374/46.
XX Novel maytansinoid produced by bacterial host cell transformed with
PT expression vector comprising open reading frame from ansamitocin gene
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX Disclosure; Page 102; 160pp; English.
XX The invention relates to maytansinoid produced by bacterial host cell
CC transformed with expression vector comprising open reading frame from
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC useful as a potent antitumor agent. The present sequence is A. pretiosum
CC ansamitocin gene cluster II Aem46 gene
XX Sequence 795 BP; 63 A; 306 C; 338 G; 88 T; 0 U; 0 Other;
SQ Query Match 9.3%; Score 48.6; DB 7; Length 795;
Best Local Similarity 47.3%; Pred. No. 3.9;
Matches 219; Conservative 0; Mismatches 234; Indels 10; Gaps 2;
QY 49 GCACCTCCGGGGGTGACAGGTGACACAGTGGAGTTGAGCAGCAGTCTATCGCG 108
DB 661 GCGGTTCGCGCGCGGTGCGAGCTCGCGCGATCTGTTCCAGTAGCGCCGACG 602
QY 109 CCGTCCGCGCATGCGGAAACACACCGGCGCAGGATCACGTGACGACGCGGATGCGCGCT 168
DB 601 CGCGCGGAGCGACGACCGCGCCCGCGGCGACGGCCCTGCGCGGACGCGGCT 542
QY 169 CGATCTGACGCTCGCGCGAGCTCGATCTGTCGCGCGGTCTCCACACAGGGGAACT 228
DB 541 CGGTGCGCGCGCGCGGTGAGCGGTCTGCCCGAGCTTGCACGAGTGGAAACGGAACG 482
QY 229 GCGTCCGTGCGACGCGCGACCGCCAGCGCGGACAGCTCGTCCAGGTGCGGTCAGCGACT 288
DB 481 CGGGCTGCGCGGTTCGCGGCGACGCGGACAGTTGCGCCAGTGG-----CGCGCA 428
QY 289 CTCGCGGTGCGGAGCGGAAACGCGACGCTCAGATCCCTGTAGTGCATCGGCTCAG 348
DB 427 CGGTGCGCTGCGCGCGCGCGCGCGCGCGGTCTCTCGCCCGACGCGCGCGCGCGCG 368
QY 349 TCGCGTCTGCTCCCTTGGCTGGGAGGATAGCGGTTACAGACGAGCGGACCGCGGG 408
DB 367 CGACGTCCAGCGCGGATTCGCTGGAGCGCGTCCGACCTCGACAGCGGCTCGACCGGAC 308
QY 409 GCGGGGCGGGGCGGCTTCAGCCGATCCGTCGATGACGAGCGGTGCGGGGTGCGGGCG 468
DB 307 CGCGCGGCAAGCGCTCGCGCGCG-----GGTCGACGCGGAGTCTCGCGGCGCGCGCTCG 252
QY 469 GTCGCGGAGATCGGTACCGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCG 511
DB 251 GCGCGGACGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 209
RESULT 27
ACA40900/C
ID ACA40900 standard; DNA; 1797 BP.
XX ACA40900;
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #22557.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Mycobacterium tuberculosis.
XX WO200277183-A2.
XX 03-OCT-2002.
```



[illegible][illegible]

Qy	417	GGGGGGCGGTTCACGCCGATCCGCTCGATCACACGCGGCTCGGGTCCGGGCGGTCCGGCGA	476
Dd	25449	GGTGGGGCCGTTGGCGGCACCGGCGGGG-CGCGCGGCGCGCCGAGCTGCTGTTCCGGCC	
Qy	477	GATCCGTACCGGCCGAGACCGCTTCGGCCAGCGCCGCGG	515
Dd	25508	CGCGGTGCGGGTGGGGCGGGCACCGACGCGGGGCGCGG	25546

RESULT 31  
ACA27331/c  
ID ACA27331 standard; DNA; 1689 BP.  
XX  
XX ACA27331;  
XX  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Prokaryotic essential gene #8988.  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX  
XX Bordetella pertussis.  
XX  
XX OS  
XX WQ200277183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0363699P.

XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR WPI; 2003-029926/02.  
XX DR P-PSDB; ABU23461.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 14; SEQ ID NO 15201; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 6213 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
XX CC encoding a polypeptide whose expression is inhibited by the antisense  
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX CC polypeptide or its fragment whose expression is inhibited by the  
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX CC proliferation or the activity of a gene in an operon required for  
XX CC proliferation; (7) identifying a compound that influences the activity of  
XX CC the gene product or that has an activity against a biological pathway  
XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
XX CC identifying a gene required for cellular proliferation or the biological  
XX CC pathway in which a proliferation-required gene or its gene product lies  
XX CC or a gene on which the test compound that inhibits proliferation of an  
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX CC compound's activity; (11) a culture comprising strains in which the gene  
XX CC product is overexpressed or underexpressed; (12) determining the extent  
XX CC to which each of the strains is present in a culture or collection of  
XX CC strains; or (13) identifying the target of a compound that inhibits the  
XX CC proliferation of an organism. The antisense nucleic acids are useful for  
XX CC identifying proteins or screening for homologous nucleic acids required  
XX CC for cellular proliferation to isolate candidate molecules for rational  
XX CC drug discovery programs, or for screening homologous nucleic acids  
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
XX CC prokaryotic essential genes. Note: The sequence data for this patent did  
XX CC not form part of the printed specification, but was obtained in  
XX CC electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 1689 BP; 221 A; 626 C; 556 G; 286 T; 0 U; 0 Other;  
Query Match 9.3%; Score 48.4; DB 7; Length 1689;  
Best Local Similarity 53.8%; Pred. No. 4;  
Matches 100; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
Qy 9 ACAGACGACCGGTGGTGGCCCGCCGCGCCATCAGCACTGCCACTCGCGGGGTGACC 68  
Db 1391 AAGAACAGCGCCCTTCGACCGCGTAGACGACGTCGTCGCGCAGCGCCGTAGCCG 1332  
Qy 69 AGGTGACAGGTGGGAGTTGAGCAGCGAGCTCATCGCCGCTCGCGCGCATGCCGAAA 128  
Db 1331 CGGTGGCCCACTGTGTCGCCAGCTCGCGCGGCTCTGCGGCTGCACTTCGACGCTGGAC 1272  
Qy 129 CACGGGCGCAGATCAGCTGACGACGCGGATCGCGCTCGATCTCGACGCTGGCCGC 198  
Db 1271 GCCATGCGCCGAGGAGTGCAGATGCGGATCACCAGCGGATGTTGACCGCCACCACC 1212  
Qy 189 AGCTCG 194  
Db 1211 AGGTCG 1206

XX ACA38077;  
XX 19-JUN-2003 (first entry)  
XX DE Prokaryotic essential gene #19734.  
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX KW drug design; gene.  
XX OS Mycobacterium avium.  
XX XX WO200277183-A2.  
XX PD 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX P-PSDB; ABU34207.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 25947; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
XX prokaryotic essential genes. Note: The sequence data for this patent did  
XX not form part of the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 3669 BP; 540 A; 1243 C; 1388 G; 498 T; 0 U; 0 Other;  
Query Match 9.3%; Score 48.4; DB 7; Length 3669;

RESULT 32  
ACA38077/c  
ID ACA38077 standard; DNA; 3669 BP.

```
Best Local Similarity 46.0%; Pred. No. 3.7;
Matches 241; Conservative 0; Mismatches 276; Indels 7; Gaps 2;

QY 3 TGCAGCAGAACGACCGGTGTCGCCCGCGCCATCAACACTGCCACTCCGCGGG 62
DB 3599 TGCAGCGCGGTACACAGGTTCGCCCACTATGACCAACTCGACGTCACGACGCGCGC 3540
QY 63 GTGACCAAGTTCGACAGGTGGAGTTGAGCAGCCAGCTCATCCGCGCTCGCGGCGCATG 122
DB 3539 GCGCCCAACCCCGCGCGCGGGATGGGAACCTGCAACGTCGACGTCGTCGCGCGGTAG 3480
QY 123 CCAGAAACACCCGGGCGCAGGATCACGTGACGACGCGCGCTCGATCTCGACGGTC 182
DB 3479 CTGCGCGCGGATACATCCGCTTCAGCGCGCACCTGGGCGCGAATCGGSCAGGTGATCGGC 3420
QY 183 GCGCGCAGCTCGATCTGTCGCCCGGTCCACACGAGGGGAACTGGCTCGGTGGCAGC 242
DB 3419 GAGAGCCGCGAGGTGGCGCGCGCGCGACACCTCGGTGATGTCGCGCGCGCGCGCAC 3360
QY 243 GGCAGCCCGACCGCGGACAGCTCGTCAGGCTGGCGTCACGGACCTCTCGCGGTTCGGGA 302
DB 3359 AGCAGCGCGACCGCGCGCGCGCACCGCACCGCGAGGCGCGTTCGGGACGCGCCGCTAG 3300
QY 303 CGGGAACCGCGCAGGTTCAGATCCCTGTGTCGTCGATCGGCTCAGTCCGTCGTCGCC 362
DB 3299 CGGTGCGAGGTTCCTCGACACCGC-GTCGATCTCGTTCGAGCGCGCGCGCGCAG 3241
QY 363 C-----TTGGCCCTGGAGGATAGCGTTTACGACGAGCGGACACGCGCGCGCGGC 416
DB 3240 CCGCGGATACGCTTCCACCGCACCGCATGCTGGCGATGTAGTCCGCGCGCAGGTGCGC 3181
QY 417 GGGGGCGGTTTCAGCGGATTCGCTCGATGACGCGGCTGCGGGGTGCGGGCGGTTCGCGCA 476
DB 3180 GTCCACCGCGCAGGTTCGATCCGACGTCCTTGGGCTCCTCGCGGTGTCGACGCTCTGGCC 3121
QY 477 GATCGTACCGCCCGGACCGCTCGGCGCGCGCGCGCGCGGATCC 520
DB 3120 GTCCGCGCGCGCGGTACGCGCTCGAGCGGCTCGCGCGCTCGCCACCGCC 3077

RESULT 33
ABQ78872/c
ID ABQ78872 standard; DNA; 90600 BP.
XX AC ABQ78872;
XX DT 23-OCT-2002 (first entry)
XX DE S. roseosporus daptomycin biosynthetic gene cluster 90kb region.
XX KW Daptomycin biosynthetic gene cluster; thioesterase; antibacterial;
XX KW fungicide; virucide; antiparasitic; immunomodulator; antilipemic;
XX KW cytosstatic; gene therapy; antimitotic; immunomodulatory; siderophore;
XX KW anti-cholesterolemic; agrochemical; gene; ds.
XX OS Streptomyces roseosporus.
XX PN WO200259322-A2.
XX PD 01-AUG-2002.
XX PF 17-OCT-2001; 2001WO-US032354.
XX PR 17-OCT-2000; 2000US-0240879P.
XX PR 28-FEB-2001; 2001US-0272207P.
XX PR 06-AUG-2001; 2001US-0310385P.
XX XX (MIAO/) MIAO V P W.
XX XX (BRIA/) BRIAN P.
XX XX (BALT/) BALTZ R H.
XX XX (SILV/) SILVA C J.
XX PA Miao VPW, Brian P, Baltz RH, Silva CJ;
XX PI
```

```
XX WPI; 2002-599794/64.
XX Isolated nucleic acid molecule from a bacterial daptomycin biosynthetic
PT gene cluster encoding a thioesterase or thioesterase domain, useful for
PT generating novel linear and cyclic peptides, and products in a cell.
XX Claim 7; Page 142-165; 227pp; English.
XX The invention relates to a novel isolated nucleic acid molecule
CC comprising a sequence that encodes a thioesterase or thioesterase domain,
CC derived from a bacterial daptomycin biosynthetic gene cluster. The
CC proteins of the invention have antibacterial, fungicide, virucide,
CC antiparasitic, immunomodulator, antilipemic, and cytostatic activity. The
CC polynucleotides may have a use in gene therapy. The compositions and
CC methods of the present invention are useful for generating novel linear
CC and cyclic peptides and improving yield of a product in a cell expressing
CC an daptomycin non-ribosomal peptide synthetase (NRPS) to be used as new
CC compounds or in producing new compounds, such as antibiotics,
CC antifungals, antivirals, antiparasitics, antimitotics, antitumour agents,
CC immunomodulatory agents, anti-cholesterolemic agents, siderophores,
CC agrochemicals and cytostatics. The sequence represents the 90kb region of
CC the S. roseosporus daptomycin biosynthetic gene cluster
XX Sequence 90600 BP; 12671 A; 32312 C; 31571 G; 14046 T; 0 U; 0 Other;
SQ Best Local Similarity 9.3%; Score 48.2; DB 6; Length 90600;
Matches 104; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
Query Match
QY 89 GAGCAGCCAGCTCATGCGCGCTCGATCTCGACGCTCGCGCGCGCAGCTCGATCTCGTCCGCGG 148
DB 78445 GAGCAGGAACCCGACGAGCGCGGCGGAGCGTTCAGATCGAAGCGCGCCACCGCTGTCGTG 78386
QY 149 CAGCAGCGGCGATGCGCGCTCGATCTCGACGCTCGCGCGCGCAGCTCGATCTCGTCCGCGG 208
DB 78385 CAGCAGCGGCTTCGATCTCGGCTTCTGCTCGCGCGCTCGAGCTTCGCGACGCTCGGTG 78326
QY 209 CTCACACACAGGGGAACTGGCTCGGTGGCAGCGGAGCCCGCAGCGCGGACAGTCTGTC 268
DB 78325 CTCACAGCGCACCGCAACCTCGCGGGGATGATCTGCACGGGCGCTGCAGCGCGCTCGTA 78265
QY 269 CAGGCTGGCGTCACGGA 285
DB 78265 CCGGAAGCGCGCGCGCA 78249

RESULT 34
ADB68842
ID ADB68842 standard; DNA; 536 BP.
XX AC ADB68842;
XX DT 04-DEC-2003 (first entry)
XX DE Minorily luxI consensus sequence DNA 14.
XX DE quorum sensing; lux homologue; luxI; ds.
XX OS Unidentified.
XX PN WO2003057902-A2.
XX PD 17-JUL-2003.
XX PF 08-JAN-2003; 2003WO-US000479.
XX PR 08-JAN-2002; 2002US-0346531P.
XX PR 07-JAN-2003; 2003US-00338110.
XX XX (FRAU ) FRAUNHOFER USA INC.
XX XX Fuhrmann JJ, Romesser JA;
XX PI
```



```
XX
DT 26-MAR-2002 (first entry)
DE Human herpesvirus 2 complete DNA genome.
KW Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
KW antibacterial; fungicide; protozoicide; antirheumatic; antiinflammatory;
KW anarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW vasculitis; ds.
XX
OS Human herpesvirus 2.
XX
XX WO200176643-A1.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-US011372.
XX
XX 07-APR-2000; 2000US-0195680P.
XX
XX (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
XX Orson FM, Kinsey BM, Bhogal BS;
XX
XX WPI; 2002-066308/09.
XX
XX Composition for oral delivery of vaccines, comprises expression vector
XX containing antigenic genomic sequence, bound to aggregated protein-
XX polycationic polymer conjugate or suspension.
XX
XX Disclosure; Page 90-132; 145pp; English.
XX
XX The invention relates to a composition comprising an expression vector
XX bound to an aggregated protein-polycationic polymer conjugate or
XX suspension. The expression vector contains a promoter polynucleotide
XX sequence operatively linked to a polynucleotide sequence encoding an
XX antigen which is a fragment of a gene or genome associated with an
XX infectious disease, cancer and autoimmune disease such as rheumatoid
XX arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
XX consisting of bacterium, fungus, protozoa and virus such as human
XX immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
XX virus (HCV), influenza and respiratory syncytial virus (RSV), and
XX optionally comprising a nucleotide sequence encoding a cytokine (or a
XX cytokine expression vector), is useful for inducing an immune response
XX (systemic and/or mucosal) in an organism. The cytokine expression vector
XX contains a sequence for granulocyte macrophage-colony stimulating factor
XX (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
XX the antigen and the cytokine are under transcriptional control of same or
XX different promoter polynucleotide sequences. The expression vector, as a
XX DNA vaccine is useful for treating a condition in an organism. The
XX present sequence is human herpesvirus 2 complete DNA genome related to
XX the invention
XX
XX Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 U; 0 Other;
XX
XX Query Match 9.2%; Score 48; DB 6; Length 154746;
XX Best Local Similarity 46.8%; Pred. No. 3.2;
XX Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;
XX
XX 23 GTGCGCCGCGCCGCCATCACGAACCTGCCACTCCGCGGGGTGACCCAGGTCCACCAAGTGTG 82
XX 129000 GTTGCCCGCCAGCGCGCGTGTGCGGCGCCGCGAGACCGTTCGCCAGGCGCCAGCAG 129059
XX
XX 83 GAGATTGAGCAGCCAGCTATCCCGCGCTGCGCGCGCGCATGCCGAACACCGGCGCAGGAT 142
XX 129060 GCAGGACAGCCCGCGTGTGCGCGGACCACTCCGCGGGGGGGCGCCCGCGCGCGGCC 129119
XX
XX 143 CAGGTGAGCAGCCGATGCGCGCCCTCGATCTCGACGTCGCGCGCGCGAGTCTCGTTC 202
XX 129120 CCGGGCCAGGTCTCTGCGCGCGAGCGCGCGGTAGAGAT-----CACCACGCGCAGTTC 129173
XX
XX 203 GCGCGGCTCCACACACAGGGGAAACTGGTCTGGTGTGACGCGGACGCCCGCGGACAG 262
XX
```

```
Db 129174 CTCGGGTCCGGGCACCTGGGCGCATCCAGGCCGCCGCGCGCGCGCGCGCGCGCGCG 129233
Qy 263 CTGTCAGGCTGGCGTCAACGACCTCTCGCGGGTGGGAGCGGAAACCGCGCACGGCTCA 322
Db 129234 CAGCGGCGCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129293
Qy 323 GATCCCTGTAGTCGCTCATCGGCTCAGTCCCGGTGCTGCCCTTGGCTTGGGAGATAGCG 382
Db 129294 CGAGGCCAGCGCGCGCGCGGTCTGCAACATGAGGGCGCGCGCGCGCGCGCGCGCG 129353
Qy 383 GTTCACGACGAGCGGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442
Db 129354 GTGCTCGTGTGCTCGGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129413
Qy 443 TGACCA-----CGGCTGCGGGGTGCGGGCGGTGCGCGCGAGATCCGTACCGCGCGCGCG 496
Db 129414 GGGCGCGCGGTGTGTTGGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129473
Qy 497 CCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516
Db 129474 GCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129493

RESULT 37
AAD25519/C
ID AAD25519 standard; DNA; 154746 BP.
XX
XX AAD25519;
AC
XX
DT 26-MAR-2002 (first entry)
XX
DE Human herpesvirus 2 complete DNA genome.
XX
KW Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
KW antibacterial; fungicide; protozoicide; antirheumatic; antiinflammatory;
KW anarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW vasculitis; ds.
XX
OS Human herpesvirus 2.
XX
XX WO200176643-A1.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-US011372.
XX
XX 07-APR-2000; 2000US-0195680P.
XX
XX (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
XX Orson FM, Kinsey BM, Bhogal BS;
XX
XX WPI; 2002-066308/09.
XX
XX Composition for oral delivery of vaccines, comprises expression vector
XX containing antigenic genomic sequence, bound to aggregated protein-
XX polycationic polymer conjugate or suspension.
XX
XX Disclosure; Page 90-132; 145pp; English.
XX
XX The invention relates to a composition comprising an expression vector
XX bound to an aggregated protein-polycationic polymer conjugate or
XX suspension. The expression vector contains a promoter polynucleotide
XX sequence operatively linked to a polynucleotide sequence encoding an
XX antigen which is a fragment of a gene or genome associated with an
XX infectious disease, cancer and autoimmune disease such as rheumatoid
XX arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
XX consisting of bacterium, fungus, protozoa and virus such as human
XX immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
XX virus (HCV), influenza and respiratory syncytial virus (RSV), and
XX optionally comprising a nucleotide sequence encoding a cytokine (or a
XX cytokine expression vector), is useful for inducing an immune response
XX (systemic and/or mucosal) in an organism. The cytokine expression vector
XX contains a sequence for granulocyte macrophage-colony stimulating factor
XX (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
XX the antigen and the cytokine are under transcriptional control of same or
XX different promoter polynucleotide sequences. The expression vector, as a
XX DNA vaccine is useful for treating a condition in an organism. The
XX present sequence is human herpesvirus 2 complete DNA genome related to
XX the invention
XX
XX Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 U; 0 Other;
XX
XX Query Match 9.2%; Score 48; DB 6; Length 154746;
XX Best Local Similarity 46.8%; Pred. No. 3.2;
XX Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;
XX
XX 23 GTGCGCCGCGCCGCCATCACGAACCTGCCACTCCGCGGGGTGACCCAGGTCCACCAAGTGTG 82
XX 129000 GTTGCCCGCCAGCGCGCGTGTGCGGCGCCGCGAGACCGTTCGCCAGGCGCCAGCAG 129059
XX
XX 83 GAGATTGAGCAGCCAGCTATCCCGCGCTGCGCGCGCGCATGCCGAACACCGGCGCAGGAT 142
XX 129060 GCAGGACAGCCCGCGTGTGCGCGGACCACTCCGCGGGGGGGCGCCCGCGCGCGGCC 129119
XX
XX 143 CAGGTGAGCAGCCGATGCGCGCCCTCGATCTCGACGTCGCGCGCGCGAGTCTCGTTC 202
XX 129120 CCGGGCCAGGTCTCTGCGCGCGAGCGCGCGGTAGAGAT-----CACCACGCGCAGTTC 129173
XX
XX 203 GCGCGGCTCCACACACAGGGGAAACTGGTCTGGTGTGACGCGGACGCCCGCGGACAG 262
XX
```



CC cytokine expression vector), is useful for inducing an immune response  
CC (systemic and/or mucosal) in an organism. The cytokine expression vector  
CC contains a sequence for granulocyte macrophage-colony stimulating factor  
CC (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding  
CC the antigen and the cytokine are under transcriptional control of same or  
CC different promoter polynucleotide sequences. The expression vector, as a  
CC DNA vaccine is useful for treating a condition in an organism. The  
CC present sequence is human herpesvirus 2 complete DNA genome related to  
CC the invention  
XX  
SQ Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 U; 0 Other;  
Query Match 9.2%; Score 48; DB 6; Length 154746;  
Best Local Similarity 46.6%; Pred. No. 3.2;  
Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;  
QY 23 GTGCGCCCGCGCCCATCACTGCGGCTCGGCGGTTGACCGAGTCGACCGAGGTG 82  
Db 152742 GTTGCCCGCCGAGCGCGCCGTGGCGGCGCCGACAGCGCGTTGCCAGGCGCCGACGAG 152683  
QY 83 CGAGTTGAGCAGCAGCTCATCCCGCTGCGCGGCGATCCCGAAGACCGGGCCAGGAT 142  
Db 152682 GCAGGACAGCGCGCGCGCTGCGCGGACCACTCCGCGGGGGCGCCCGCGCGCGCC 152623  
QY 143 CACGTGACGACGCGGATGCGCGCTCGATCTCGACGCTCGGCGCGCAGTCGATCTCGTC 202  
Db 152622 CGCGGCCAGTCTCTCGCGCGCGCGCGCGCGAGTAGAGAT-----CACCAAGCGCACGTC 152569  
QY 203 GCGCGGTCCACACAGGAGGAACTGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGAG 262  
Db 152568 CTCCGGGTGCGGCGACCTGCGCGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCG 152509  
QY 263 CTGCTTCAGGCTGCGGTCACGAGACCTCTCGCGGCTCGGAGCGGAAACCGCGCGCTCA 322  
Db 152508 CAGCGGCGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152449  
QY 323 GATCCCTGTGCTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 382  
Db 152448 CGAGCG 152389  
QY 383 GTTACGACGAGCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442  
Db 152388 GTGCTCGGTGAGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152329  
QY 443 TGACCA-----GCGGCTGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGG 496  
Db 152328 GCGGCGCGCGCGTGTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152269  
QY 497 CCTCGGCG 516  
Db 152268 GCGCTCGGCG 152249  
RESULT 38  
ID ABZ78139  
XX ABZ78139 standard; cDNA; 1679 BP.  
AC ABZ78139;  
XX  
DT 19-MAY-2003 (first entry)  
XX  
DE Human cancer-related coding sequence, 187P3P2.  
XX  
KW Human; cytostatic; vaccine; cancer; immune response; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200281921-A2.  
XX  
PD 24-OCT-2002.  
XX  
PF 10-APR-2002; 2002WO-US011654.  
XX

PR 10-APR-2001; 2001US-0282739P.  
PR 10-APR-2001; 2001US-0283112P.  
PR 25-APR-2001; 2001US-028630P.  
XX  
PA (AGEN-) AGENSYS INC.  
XX  
PI Jakobovits A, Challita-Eid PM, Faris M, Ge W, Hubert RS;  
PI Morrison K, Morrison RK, Raitano AB;  
XX  
DR WPI; 2003-075555/07.  
DR P-PSDB; ABR01808.  
XX  
XX New composition comprising a substance that modulates the structure of  
PT proteins and polynucleotides, useful for therapeutic, prognostic and  
PT diagnostic reagents for eliciting cellular or humoral immune response in  
PT cancer patients.  
XX  
XX Claim 26; Fig 2Q; 1021pp; English.  
XX  
XX The present invention relates to novel human cancer-related genes and  
CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and  
CC proteins are useful for eliciting a humoral or cellular immune response.  
CC The genes are useful as probes and primers for the amplification and/or  
CC detection of genes, mRNAs or their fragments, as reagents for the  
CC diagnosis and/or prognosis of cancer, as coding sequences capable of  
CC inhibiting the expression of the protein, as tools for modulating or  
CC as therapeutic agents. The proteins and peptides are useful as  
CC therapeutic, prognostic and diagnostic reagents for cancer. 74P3B3 maps  
CC to chromosome 6p21.3-p22, 83P488 maps to chromosome 15q25.1, 109P1D4 maps  
CC to chromosome 11q13.3-q13.5, 154P2A8 maps to chromosome 2p13.1-p12,  
CC maps to chromosome Xp22.3, 156P5C12 maps to chromosome 2p13.1-p12,  
CC 179P3G7 maps to chromosome 7p15.1, 161P2B7A maps to chromosome 3q25-q26,  
CC 19P13.1, 184P3G10 maps to chromosome 12q12-q13, 184P3C10B maps to chromosome  
CC 18p11.22, 185P3C2 maps to chromosome 2p12, 185P2C9 maps to chromosome  
CC 7q21.3-q22.1, 187P3P2 maps to chromosome 17q21, 186P1H9 maps to chromosome  
CC chromosome 22q13.2-q13.31  
XX  
SQ Sequence 1679 BP; 255 A; 647 C; 615 G; 162 T; 0 U; 0 Other;  
Query Match 9.2%; Score 47.6; DB 7; Length 1679;  
Best Local Similarity 45.9%; Pred. No. 5.6;  
Matches 199; Conservative 0; Mismatches 234; Indels 1; Gaps 1;  
QY 76 CCAGTGGAGTTGAGCAGCAGCTCATCGCGCTGCGCGCGCATGCGGAACACCGCG 135  
Db 458 CCAGCAGCCACCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 517  
QY 136 CCAGGATCAGTCAGCAGCAGCGCGGATCGCGCGCTCGATCTCGACGCTCGCGCGCGCTCGA 195  
Db 518 CGGCGCGAGCAGCTGACCGCGCGCAGCGCTGACACCGCGCGCGCGCGCGCGCGCGCG 577  
QY 196 TCTCGTCGCGCGCTCCACACACAGGGGAAACTGGCTCGGTGGGAGCGCGAGCCCGAGCC 255  
Db 578 ACCCGCTGC 637  
QY 256 GGGACAGCTCTCCAGGCTGGGTCAACGACCTCTCGCGGGTCCGAGCGGAAACGCGCA 315  
Db 638 CGCAGCGCGCAGCG 696  
QY 316 CGGCTCAGATCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 375  
Db 697 AGCAGCG 756  
QY 376 GATAGCGGTTACAGCAGCAGCGGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 435  
Db 757 TGCTGAGCG 816  
QY 436 CGCTCGATGACACAGCGCGCTGCGGGGTGCGGGGTGCGGGGTGCGGGGTGCGGGGTGCG 495  
Db 817 GCTTGTGTCACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 876

XX  
SQ Sequence 4523 BP; 723 A; 1506 C; 1619 G; 675 T; 0 U; 0 Other;  
Query Match 9.2%; Score 47.6; DB 9; Length 4523;  
Best Local Similarity 53.2%; Pred. No. 5.1;  
Matches 101; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 325 TCCTGTGTCATCGCTCAGTCCGGTCTGCTCCCTTGGCTGGGAGATAGCGGT 384  
DB 3216 TCGTGTAGGCCCGCCGAGCGCGCGGACCGGGCCCGGCGGACCCGCG 3157  
QY 385 TCACGACGAGCGGCACACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 444  
DB 3156 GGGGAGCGCGCGCGCCCTCCCGCGCTCCCGCGCGCGCGCGCGAGAGGGGG 3097  
QY 445 ACCAGCGGCTCGCGGGTTCGGGGCGGTCGGCGAGATCCGTACCGCCCGGACCGCTCGGCC 504  
DB 3096 AACGGGGGGGACGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 3037  
QY 505 AGCGCCGCGG 514  
DB 3036 GACGCGCGG 3027  
RESULT 40  
ADE60229/C  
ID ADE60229 standard; DNA; 4523 BP.  
XX  
AC ADE60229;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human gene M11167, SEQ ID NO 6135.  
XX  
KW Human; ds; gene; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
DR WPI; 2003-268312/26.  
DR GENBANK; M11167.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates

QY 496 GCGTCGGCCAGCGC 509  
DB 877 ACCACACACACAC 890  
RESULT 39  
ADE60227/C  
ID ADE60227 standard; DNA; 4523 BP.  
XX  
AC ADE60227;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human gene M11167, SEQ ID NO 6133.  
XX  
KW Human; ds; gene; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
DR WPI; 2003-268312/26.  
DR GENBANK; M11167.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the  
CC specification) which encodes one of the polypeptides of the invention  
CC which is differentially expressed during pain. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.



XX Homo sapiens;  
XX WO2003016475-A2.  
XX 27-FEB-2003.  
XX  
XX 14-AUG-2002; 2002WO-US025765.  
XX  
XX 14-AUG-2001; 2001US-0312147P.  
XX 01-NOV-2001; 2001US-0346382P.  
XX 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEO) GEN HOSPITAL CORP.  
XX (PABB) BAYER AG.  
XX  
XX Woolf C, D'urso D, Befort K, Costigan M;  
XX WPI; 2003-268312/26.  
XX GENBANK; M11167.  
XX  
XX New composition comprising two or more isolated polypeptides, useful for  
XX preparing a medicament for treating pain in an animal.  
XX  
XX Claim 1; Page; 1017pp; English.  
XX  
XX The invention discloses a composition comprising two or more isolated rat  
XX or human polynucleotides or a polynucleotide which represents a fragment,  
XX derivative or allelic variation of the nucleic acid sequence. Also  
XX claimed are a vector comprising the novel polynucleotide, a host cell  
XX comprising the vector, a method for identifying a nucleotide sequence  
XX which is differentially regulated in an animal subjected to pain and a  
XX kit to perform the method, an array, a method for identifying an agent  
XX that increases or decreases the expression of the polynucleotide sequence  
XX that is differentially expressed in neuronal tissue of a first animal  
XX subjected to pain, a method for identifying a compound which regulates  
XX the expression of a polynucleotide sequence which is differentially  
XX expressed in an animal subjected to pain, a method for identifying a  
XX compound that regulates the activity of one or more of the  
XX polynucleotides, a method for producing a pharmaceutical composition, a  
XX method for identifying a compound or small molecule that regulates the  
XX activity in an animal of one or more of the polypeptides given in the  
XX specification, a method for identifying a compound useful in treating  
XX pain and a pharmaceutical composition comprising the one or more  
XX polypeptides or their antibodies. The polynucleotide or the compound that  
XX modulates its activity is useful for preparing a medicament for treating  
XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
XX therapy). The sequence presented is a human DNA (shown in Table 2 of the  
XX specification) which encodes one of the polypeptides of the invention  
XX which is differentially expressed during pain. Note: The sequence data  
XX for this patent did not form part of the printed specification, but was  
XX obtained in electronic form directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 9.2%; Score 47.6; DB 9; Length 4523;  
Best Local Similarity 53.2%; Fred. No. 5.1;  
Matches 101; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
XX 325 TCCTGTGTCAGTCGCTCAGTCGCGGTGCTGCCCTTGGCTGGGAGATAGCGGT 384  
DB 3216 TCCTGTGTCAGTCGCGGCGAGGCGCGCGGACCGGCGCGGCGGCGGCGGCGG 3157  
XX 385 TCACACAGCGGACACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 444  
DB 3156 GGGGACGG 3097  
XX 445 ACCACGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCG 504  
DB 3096 AACGGGGGGGACGCGGCGGCGGCGGTAGCGGCGGCGGCGAAACGCGCGCGCGCGCGCG 3037

QY 505 AGCGCGCGCG 514  
DB 3036 GACGCGCGCG 3027  
RESULT 43  
ADD45822/c  
ID ADD45822 standard; DNA; 4523 BP.  
XX  
XX AC ADD45822;  
XX DT 29-JAN-2004 (first entry)  
XX  
XX Human gene M11167, SEQ ID NO 11491.  
XX  
XX Human; ds; gene; pain; neuronal tissue; gene therapy;  
XX spinal segmental nerve injury; chronic constriction injury; CCI;  
XX spared nerve injury; SNI; Chung.  
XX  
XX Homo sapiens.  
XX  
XX WO2003016475-A2.  
XX  
XX 27-FEB-2003.  
XX  
XX 14-AUG-2002; 2002WO-US025765.  
XX  
XX 14-AUG-2001; 2001US-0312147P.  
XX 01-NOV-2001; 2001US-0346382P.  
XX 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEO) GEN HOSPITAL CORP.  
XX (PABB) BAYER AG.  
XX  
XX Woolf C, D'urso D, Befort K, Costigan M;  
XX WPI; 2003-268312/26.  
XX GENBANK; M11167.  
XX  
XX New composition comprising two or more isolated polypeptides, useful for  
XX preparing a medicament for treating pain in an animal.  
XX  
XX Claim 1; Page; 1017pp; English.  
XX  
XX The invention discloses a composition comprising two or more isolated rat  
XX or human polynucleotides or a polynucleotide which represents a fragment,  
XX derivative or allelic variation of the nucleic acid sequence. Also  
XX claimed are a vector comprising the novel polynucleotide, a host cell  
XX comprising the vector, a method for identifying a nucleotide sequence  
XX which is differentially regulated in an animal subjected to pain and a  
XX kit to perform the method, an array, a method for identifying an agent  
XX that increases or decreases the expression of the polynucleotide sequence  
XX that is differentially expressed in neuronal tissue of a first animal  
XX subjected to pain, a method for identifying a compound which regulates  
XX the expression of a polynucleotide sequence which is differentially  
XX expressed in an animal subjected to pain, a method for identifying a  
XX compound that regulates the activity of one or more of the  
XX polynucleotides, a method for producing a pharmaceutical composition, a  
XX method for identifying a compound or small molecule that regulates the  
XX activity in an animal of one or more of the polypeptides given in the  
XX specification, a method for identifying a compound useful in treating  
XX pain and a pharmaceutical composition comprising the one or more  
XX polypeptides or their antibodies. The polynucleotide or the compound that  
XX modulates its activity is useful for preparing a medicament for treating  
XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
XX therapy). The sequence presented is a human DNA (shown in Table 2 of the  
XX specification) which encodes one of the polypeptides of the invention  
XX which is differentially expressed during pain. Note: The sequence data  
XX for this patent did not form part of the printed specification, but was  
XX obtained in electronic form directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.

FT	/product= "Protein of ORF 11"	
FT	15880. .19035	
FT	*tag= l	
FT	/product= "Protein of ORF 12"	
FT	19032. .39713	
FT	*tag= m	
FT	/product= "Protein of ORF 13"	
FT	39713. .65800	
FT	/tag= n	
FT	/product= "Protein of ORF 14"	
FT	65826. .66530	
FT	*tag= o	
FT	/product= "Protein of ORF 15"	
FT	66546. .66730	
FT	*tag= p	
FT	/product= "Protein of ORF 16"	
FT	67384. .70059	
FT	/tag= q	
FT	/product= "Protein of ORF 17"	
FT	70099. .70662	
FT	*tag= r	
FT	/product= "Protein of ORF 18"	
FT	70659. .71906	
FT	*tag= s	
FT	/product= "Protein of ORF 19"	
FT	complement(73433. .71964)	
FT	/tag= t	
FT	/product= "Protein of ORF 20"	
FT	complement(74216. .73563)	
FT	*tag= u	
FT	/product= "Protein of ORF 21"	
FT	complement(75424. .74213)	
FT	*tag= v	
FT	/product= "Protein of ORF 22"	
FT	75535. .76464	
FT	/tag= w	
FT	/product= "Protein of ORF 23"	
FT	complement(78110. .76449)	
FT	*tag= x	
FT	/product= "Protein of ORF 24"	
FT	complement(79864. .78107)	
FT	*tag= y	
FT	/product= "Protein of ORF 25"	
FT	complement(81624. .79861)	
FT	/tag= z	
FT	/product= "Protein of ORF 26"	
FT	complement(81909. .81682)	
FT	*tag= aa	
FT	/product= "Protein of ORF 27"	
FT	complement(82346. .82062)	
FT	/tag= ab	
FT	/product= "Protein of ORF 28"	
FT	83587. .84446	
FT	/tag= ac	
FT	/product= "Protein of ORF 29"	
FT	84481. .85548	
FT	*tag= ad	
FT	/product= "Protein of ORF 30"	
FT	85556. .86845	
FT	/tag= ae	
FT	/product= "Protein of ORF 31"	
FT	87372. .86803	
FT	*tag= af	
FT	/product= "Protein of ORF 32"	
FT	87494. .88420	
FT	*tag= ag	
FT	/product= "Protein of ORF 33"	
XX		
PN	WO200231155-A2.	
XX		
PD	18-APR-2002.	
XX		
PP	15-OCT-2001; 2001WO-CA001462.	

XX 13-OCT-2000; 2000US-0239924P.  
PR 12-APR-2001; 2001US-0283296P.  
PR 24-JUL-2001; 2001US-00910813.  
XX  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Farnet CM, Zazopoulos E, Staffa A;  
XX  
XX WPI; 2002-435445/46.  
DR P-PSDB; AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,  
DR AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,  
DR AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,  
DR AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,  
DR AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.  
XX  
XX Novel isolated ramoplanin biosynthetic pathway polypeptide useful for  
XX chemically modifying biological molecule that is a substrate for a  
XX polypeptide encoded by a ramoplanin biosynthesis gene cluster.  
XX  
XX Disclosure; Page 87-135; 212pp; English.  
XX  
XX The invention relates to an isolated ramoplanin biosynthetic pathway  
XX polypeptide selected from a polypeptide of open reading frames (ORF) 1-  
XX 32. The isolated polypeptides are useful for chemically modifying a  
XX biological molecule that is a substrate for a polypeptide encoded by a  
XX ramoplanin biosynthesis gene cluster, by contacting the biological  
XX molecule with the isolated polypeptide, where the polypeptide chemically  
XX modifies the biological molecule. The method comprises contacting the  
XX biological molecule with at least two different polypeptides encoded by  
XX ramoplanin ORFs 1-31. The polypeptides are useful for directing the  
XX biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated  
XX gene cluster comprising the ORFs is useful as a substrate for  
XX bioengineering of antibiotic structures. An isolated polypeptide or its  
XX encoding nucleic acid sequence is useful for generating derivatives of  
XX ramoplanin, for improving production or for producing variants of other  
XX antibiotics of the peptide class. The isolated polypeptides are useful  
XX for synthesis of ramoplanin in vivo or in vitro, as an acylation domain  
XX in conjunction with other peptide synthetase modules and allowing the  
XX incorporation of Thr into a peptide antibiotic precursor, for modifying  
XX fatty acid structure and/or enhancing fatty acid incorporation into the  
XX peptide antibiotic structure, for production of an hydroxyphenylglycine  
XX (HPG)-containing peptide antibiotic, for enhancing secretion of  
XX ramoplanin or its variants and derivatives, for enhancing uptake of  
XX precursors for ramoplanin biosynthesis, for enhancing production of  
XX ramoplanin products or its variants or derivatives, to chlorinate HPG of  
XX a peptide antibiotic precursor, and for designing specific nucleotide  
XX probes and primers for identifying and isolating putative lipdipeptide  
XX -producing microorganisms. This polynucleotide sequence represents the  
XX 88421nt genomic DNA of a ramoplanin producing Actinoplanes sp.  
XX microorganism of the invention  
XX  
XX Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 U; 0 Other;  
XX  
XX Query Match 9.1%; Score 47.4; DB 6; Length 88421;  
XX Best Local Similarity 49.8%; Pred. No. 4.3;  
XX Matches 120; Conservative 0; Mismatches 121; Indels 0; Gaps 0;  
XX  
XX 43 CGAATGCTCCATCGCGGGGTGACAGGTGACAGGTGGAGTTGAGCCAGCTCA 102  
XX 59630 CGACGCGCTGCGATCGGGATGTGCGGCGCGCGCGGCGGACAGCGTCCGA 59571  
XX  
XX 103 TCGCGCTGCGCGCGATGCGGAACACCGGCGCGAGTACGTCGACGCGGATGC 162  
XX 59570 AGCGCGCTGAGCAGCATGACAGGTGACCGCTCGACCGGCGCGAGTCCGCGG 59511  
XX  
XX 163 GCGCTCGATTCGAGGTGCGCGGAGTTCGATTCGTCGCGCGCTCCACACAGGG 222  
XX 59510 GCGGTGCACTCGCGCGGACCGCGGACCGGTGCTGCTGCTGCTGCTGCTGCTG 59451  
XX  
XX 223 GAAACTGGCTCGTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 282  
XX 59450 CCGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 59391

QY 283 G 283  
DB 59390 G 59390

## RESULT 45

ABK32842/C  
ID ABK32842 standard; DNA; 2307 BP.  
XX  
XX AC ABK32842;  
XX  
XX 23-APR-2002 (first entry)  
DT  
XX  
DE DNA encoding human homologue of MPT1 antifungal target.  
XX  
XX antifungal; fungal gene transcription; RPC34; POP3; TPA2; NAB2; MPT1;  
XX MTR2; BOS1; POP30; RSA2; SQT1; MTW1; TFB1; SPC99; BFR2; RNAL; GCD7; SK16;  
XX NTP1; LCP5; NCE103; ECO1; ORC2; CNS1; YPD1; TIM10; SRB4; Yeast; fungus;  
XX ds; gene.  
XX  
XX Homo sapiens.  
XX  
XX WO200202055-A2.  
XX  
XX 10-JAN-2002.  
XX  
XX 28-JUN-2001; 2001WO-US020592.  
XX  
XX 29-JUN-2000; 2000US-0215164P.  
XX 18-AUG-2000; 2000US-0224457P.  
XX  
XX (ANAD-) ANADYS PHARM INC.  
XX  
XX Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;  
XX Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S, Long P;  
XX Davidov E, Thompson CM;  
XX  
XX WPI; 2002-147962/19.  
XX P-PSDB; AAU82954.  
XX  
XX Screening candidate antifungal compound for interaction with essential  
XX protein, modulation of essential protein activity, binding to essential  
XX protein, by contacting protein with test compound and determining  
XX effects.  
XX  
XX Disclosure; Fig 80; 522pp; English.  
XX  
XX The invention describes a method of screening a candidate antifungal  
XX compound for interaction with essential proteins (EP) or for modulation  
XX of EP activity e.g. fungal gene transcription. The proteins tested in the  
XX invention include RPC34, POP3, TPA2, NAB2, MPT1, MTR2, BOS1, POP30, RSA2,  
XX SQT1, MTW1, TFB1, SPC98, BFR2, RNAL, GCD7, SK16, NTP1, LCP5, NCE103,  
XX ECO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from *S. cerevisiae*, *C. albicans*  
XX or more test compounds and determining the effects on the growth or  
XX viability of the culture of cells which preferably comprises fungal cells  
XX or yeast cells. Preferably the identified compounds interact with, or  
XX modulate (preferably inhibit) activity of *C. albicans* EP. The inhibitor  
XX compounds identified by the method are useful for preventing or  
XX inhibiting fungal, particularly *C. albicans* growth in culture or in a  
XX mammal. The antifungal agents interact with essential fungal elements  
XX that can be used to treat fungal infection by preventing the growth and  
XX preferentially killing the fungi, but does not inhibit the biological  
XX activity of mammalian homologues. This sequence encodes a target protein  
XX used to test the antifungal compounds, described in the method of the  
XX invention  
XX  
XX Sequence 2307 BP; 359 A; 969 C; 724 G; 255 T; 0 U; 0 Other;  
XX

Query Match 9.1%; Score 47.2; DB 6; Length 2307;  
Best Local Similarity 44.6%; Pred. No. 6.4;  
Matches 232; Conservative 0; Mismatches 283; Indels 5; Gaps 1;

QY 2 GTGACGACAAACAGCGGTGTCGCCCGCGCCCACTGCAAACTGCCACTCCGGCGG 61  
Db 1269 GGGCGTCCGGGACAGGCTCTGGGTCACTGCGCGCGCGCGCTTTGGGGCAGCCCGTGGG 1210  
QY 62 GGTGACAGAGTGCACAGAGTGGAGTTGAGCAGCAGCTCATCGCGCGCTGGCGCGGCAT 121  
Db 1209 GGTCCCGGGCGCGCGCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1150  
QY 122 GCCGAAACACCGCGCCAGGATCACGTGACAGCGCGGATGCGCGCGCTTCGACACGCT 181  
Db 1149 CCGGATGACCATGCTGCGCCCGCTGCTGCGCGCGCGCTGGCGCGCGCGCTGCGCGCG 1090  
QY 182 CGCGCGAGTGCATCTGTCGCCCGCGCTCCCAACAGCGGAACTGCTGCTGGTGGCAG 241  
Db 1089 CG 1030  
QY 242 CGGACG 301  
Db 1029 CGACG 970  
QY 302 AGCGGAAACCG 356  
Db 969 GCG 910  
QY 357 GTCCCGCTTGGCG 416  
Db 909 ATACCTTGGCG 850  
QY 417 GGGCGCGCGTTGACCGCGATCGCTGATGACAGCGCGCGCGCGCGCGCGCGCGCG 476  
Db 849 GCG 790  
QY 477 GATCGTACCG 516  
Db 789 GCG 750  
RESULT 46  
ID ABX72218/c  
XX ID ABX72218 standard; cDNA; 5571 BP.  
AC ABX72218,  
XX  
DT 03-JUN-2003 (first entry)  
XX  
DE Human NOVX polynucleotide #49.  
XX  
KW Human; NOVX; gene; ss; metabolic disorder; cardiomyopathy; diabetes; ASD;  
KW hypertension; congenital heart defect; aortic stenosis; valve disease;  
KW atrial septal defect; atrioventricular canal defect; ductus arteriosus;  
KW pulmonary stenosis; subaortic stenosis; ventricular septal defect; VSD;  
KW tuberosus sclerosis; scleroderma; atherosclerosis; infectious disease;  
KW obesity; anorexia; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; immune disorder; haematopoietic disorder;  
KW haemophilia; hypercoagulation; Crohn's disease; cancer.  
XX  
OS Homo sapiens.  
XX  
FN WO200281498-A2.  
XX  
PD 17-OCT-2002.  
XX  
FF 03-APR-2002; 2002WO-US010780.  
XX  
FF 03-APR-2001; 2001US-0281086P.  
PR 03-APR-2001; 2001US-0281136P.  
PR 05-APR-2001; 2001US-0281863P.  
PR 05-APR-2001; 2001US-0281906P.  
PR 06-APR-2001; 2001US-0282020P.  
PR 10-APR-2001; 2001US-0282930P.  
PR 10-APR-2001; 2001US-0282934P.  
PR 12-APR-2001; 2001US-0283512P.

PR 13-APR-2001; 2001US-0283710P.  
PR 17-APR-2001; 2001US-0284234P.  
PR 19-APR-2001; 2001US-0285325P.  
PR 20-APR-2001; 2001US-0285381P.  
PR 20-APR-2001; 2001US-0285609P.  
PR 23-APR-2001; 2001US-0285748P.  
PR 23-APR-2001; 2001US-0285890P.  
PR 24-APR-2001; 2001US-0286088P.  
PR 24-APR-2001; 2001US-0286252P.  
PR 27-APR-2001; 2001US-0287213P.  
PR 02-MAY-2001; 2001US-0288257P.  
PR 29-MAY-2001; 2001US-0294164P.  
PR 30-MAY-2001; 2001US-0294484P.  
PR 18-JUN-2001; 2001US-0298952P.  
PR 19-JUN-2001; 2001US-0299237P.  
PR 19-JUN-2001; 2001US-0299276P.  
PR 12-SEP-2001; 2001US-0318750P.  
PR 25-SEP-2001; 2001US-0324800P.  
PR 25-SEP-2001; 2001US-0324802P.  
PR 27-SEP-2001; 2001US-0325684P.  
PR 17-OCT-2001; 2001US-0330143P.  
PR 14-NOV-2001; 2001US-0332131P.  
PR 14-NOV-2001; 2001US-0332240P.  
PR 14-NOV-2001; 2001US-0332779P.  
PR 21-NOV-2001; 2001US-0332115P.  
PR 04-DEC-2001; 2001US-0337621P.  
PR 03-JAN-2002; 2002US-0345783P.  
PR 16-JAN-2002; 2002US-0350251P.  
PR 02-APR-2002; 2002US-00114270.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Guo X, Kekuda R, Miller CE, Malyankar UM, Spytek KA,  
PI Patutarajan M, Liu X, Gusev VV, Li L, Vernet CAM, Zerhusen BD;  
PI Gorman L, Shenoy SG, Pena CE, Smithson G, Burgess CE, Gerlach V;  
PI Padigaru M, Shimkets RA, Gangolli EA, Taupier RJ, Casman SJ, Ji W;  
PI Anderson DW, Leite MW, Rastelli L, Edinger SR, Stone DJ;  
PI Macdougall JR, Rothenberg ME, Mazur A, Millet I, Peyman JA;  
PI Ellerman K;  
XX  
DR WPI: 2003-046858/04.  
XX P-PSDB; ABUS4590.  
XX  
PT New isolated NOVX polypeptide useful for treating atherosclerosis,  
PT metabolic disorders, diabetes, obesity, infectious disease, anorexia,  
PT neurodegenerative disorders, Alzheimer's disease and cancer.  
XX  
PS Claim 17; Page 196-197; 666pp; English.  
XX  
CC The invention relates to human polypeptides, termed NOVX, and the  
CC polynucleotides encoding them. The polypeptides and polynucleotides are  
CC useful for diagnosing disease, and screening for potential therapeutic  
CC agents. The sequences are useful for treating metabolic disorders,  
CC cardiomyopathy, diabetes, hypertension, congenital heart defects, aortic  
CC stenosis, atrial septal defect (ASD), atrioventricular canal defect,  
CC ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular  
CC septal defect (VSD), valve diseases, tuberculous sclerosis, scleroderma,  
CC atherosclerosis, obesity, infectious disease, anorexia, neurodegenerative  
CC disorders, Alzheimer's disease, Parkinson's disease, immune disorders,  
CC haematopoietic disorders, haemophilia, hypercoagulation, Crohn's disease,  
CC and cancer. Sequences ABX72170-ABX72275 represent human NOVX  
CC polynucleotides of the invention  
XX  
SQ Sequence 5571 BP; 924 A; 2087 C; 1813 G; 747 T; 0 U; 0 Other;  
Query Match 9.1%; Score 47.2; DB 7; Length 5571;  
Best Local Similarity 47.1%; Pred. NO. 5.9;  
Matches 178; Conservative 0; Mismatches 199; Indels 2; Gaps 1;  
QY 8 CACGAACGACCGGTGTCGCCCGCGCCCACTGCAAACTGCCACTCCGGCGGTCAC 67  
Db 390 CACGGCGCGCGCGCGCGGTTCAGCGCGGACCTTCTGCTGTCAGTCCGGGATGCCGAC 331







```
QY 451 GCCTGCGGGTGGGGCGGTGGCGAGATCGGTACCGCCCGGACCGCTCGGCGCAGCGCC 510
Db 173 CGCGCCCGCGCAGCGCGCGCCAGCGCGCTGCTCGCCCCAGAGCGGATCGCCAGCGTC 114
QY 511 GCGCGGATC 519
Db 113 TCCGCTTC 105

RESULT 50
AAI19682_39
Continuation (40 of 45) of AAI19682 from base 3900001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI19682 Accession AAI19682
WP Fragment Name Begin End
WP AAI19682_00 1 110000
WP AAI19682_01 100001 210000
WP AAI19682_02 200001 310000
WP AAI19682_03 300001 410000
WP AAI19682_04 400001 510000
WP AAI19682_05 500001 610000
WP AAI19682_06 600001 710000
WP AAI19682_07 700001 810000
WP AAI19682_08 800001 910000
WP AAI19682_09 900001 1010000
WP AAI19682_10 1000001 1110000
WP AAI19682_11 1100001 1210000
WP AAI19682_12 1200001 1310000
WP AAI19682_13 1300001 1410000
WP AAI19682_14 1400001 1510000
WP AAI19682_15 1500001 1610000
WP AAI19682_16 1600001 1710000
WP AAI19682_17 1700001 1810000
WP AAI19682_18 1800001 1910000
WP AAI19682_19 1900001 2010000
WP AAI19682_20 2000001 2110000
WP AAI19682_21 2100001 2210000
WP AAI19682_22 2200001 2310000
WP AAI19682_23 2300001 2410000
WP AAI19682_24 2400001 2510000
WP AAI19682_25 2500001 2610000
WP AAI19682_26 2600001 2710000
WP AAI19682_27 2700001 2810000
WP AAI19682_28 2800001 2910000
WP AAI19682_29 2900001 3010000
WP AAI19682_30 3000001 3110000
WP AAI19682_31 3100001 3210000
WP AAI19682_32 3200001 3310000
WP AAI19682_33 3300001 3410000
WP AAI19682_34 3400001 3510000
WP AAI19682_35 3500001 3610000
WP AAI19682_36 3600001 3710000
WP AAI19682_37 3700001 3810000
WP AAI19682_38 3800001 3910000
WP AAI19682_39 3900001 4010000
WP AAI19682_40 4000001 4110000
WP AAI19682_41 4100001 4210000
WP AAI19682_42 4200001 4310000
WP AAI19682_43 4300001 4410000
WP AAI19682_44 4400001 4411529

Query Match 9.0%; Score 47; DB 4; Length 110000;
Best Local Similarity 45.3%; Pred. No. 4.9; Mismatches 250; Indels 1; Gaps 1;
Matches 208; Conservative 0;

QY 57 GCGCGGTGACAGGTGACAGGTGGAGTTGAGCAGCAGCTCATCGCGCGCTGGCGCC 116
Db 31301 GCGGTGATCAATGCGCCACCGCGCTGCTGGGGCGCCGTTGATCGGTGACCGGCC 31360
QY 117 GCGATCGCGAACAACACGGGCGCAGGATCAGTGCAGCAGCGGAGTCGGCGCTCGATTCG 176
Db 31361 AATGCGACACTCCGCGCGGGCCCGCGCGGCTGCTGTTCCGCAACGGCGG 31420
```

```
QY 177 ACGGTGCGCCGAGCTCGATCTCTGTCGCCCGGCTCCCAACACAGGGGAAACTGGTTCGTT 236
Db 31421 GCGGGGCGAGCCGGGGCGCCCGCCAGGCCCGCGGGCCTTGGCGGCCCGCGGATTTGTGG 31480
QY 237 GGCAGCGGCGAGCCCGCAGCCCGGACAGCTCGTCCAGCTGGCGTCAAGACCTCTCGCGGG 296
Db 31481 GGCACGCGGGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 31540
QY 297 TCGGAGCGGAAAACGCGCACCGCTCAGATCCCTGTCAGTTCGATCGGCTCAGTCCGCTC 356
Db 31541 GCGGTGGGTGGCTGTTTCGGGGTTGGCGGCGCGCGGGTGTCTGGTGGGCGCGTGGCGG 31600
QY 357 GTCCCCCTTGGCCTTGGGAGGATAGCGTTTCAAGACAGCGGCGCACACGCGCGGGGGGCG 416
Db 31601 ACCGGCGGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 31660
QY 417 GGGGGCGGTTTCAGCCGATCCGCTCGATGACAGCGGCTCGGGGTCGGGGCGGTCGGCGGA 476
Db 31661 GGTGGGGCGGTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 31719
QY 477 GATCCGTACCGCCCGGACCGGCTTCGGCCAGCGCGCGCGG 515
Db 31720 CGGCGGTGGGGTGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 31758
```

Search completed: June 27, 2004, 18:18:40  
Job time : 346.02 secs



Micromonosporineae; Micromonosporaceae; Micromonospora.

## REFERENCE

1 Hosted, T.J., Horan, A.C. and Wang, T.X.  
Evernimycin biosynthetic genes  
Patent: WO 0151639-A 1 19-JUL-2001;  
Schering Corporation (US)

## FEATURES

source

1. 109519  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

## ORIGIN

Query Match 100.0%; Score 520; DB 6; Length 109519;  
Best Local Similarity 100.0%; Pred. No. 7.3e-68;  
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGACAGCAACGACCGGTGGTTCGCCCGCGCCATCAGAACTGCCTCCGGCG 60  
Db 109000 GGTGACAGCAACGACCGGTGGTTCGCCCGCGCCATCAGAACTGCCTCCGGCG 109059  
QY 61 GGTGACAGGTGACAGGTGGAGTTGAGCAGCCAGTCTATCGCGCGTTCGGCGGCA 120  
Db 109060 GGTGACAGGTGACAGGTGGAGTTGAGCAGCCAGTCTATCGCGCGTTCGGCGGCA 109119  
QY 121 TGCCGAAACACCGCGCCAGGATCAGTGCAGCAGCGCGATCGCGCTCGATCTCCAGCG 180  
Db 109120 TGCCGAAACACCGCGCCAGGATCAGTGCAGCAGCGCGATCGCGCTCGATCTCCAGCG 109179  
QY 181 TCGCCCGCAGTCTCGTTCGCCCGCGTCCACACACCGGAAACTGCTCGTGGCA 240  
Db 109180 TCGCCCGCAGTCTCGTTCGCCCGCGTCCACACACCGGAAACTGCTCGTGGCA 109239  
QY 241 CGCGACGCGCCAGCGGAGAGTCTGTCAGGCTGGGTTCAGGACCTCTCGCGGTTCGG 300  
Db 109240 CGCGACGCGCCAGCGGAGAGTCTGTCAGGCTGGGTTCAGGACCTCTCGCGGTTCGG 109299  
QY 301 GAGCGGAAACGCGCACCGGCTCAGATCCCTGTCAGTGCATCGGCTCAGTCCCGTCTCC 360  
Db 109300 GAGCGGAAACGCGCACCGGCTCAGATCCCTGTCAGTGCATCGGCTCAGTCCCGTCTCC 109359  
QY 361 CCTTGGCTGGAGATAGGTTACAGCAGCGGACACACGCGCGCGCGGGGGGG 420  
Db 109360 CCTTGGCTGGAGATAGGTTACAGCAGCGGACACACGCGCGCGCGGGGGGGGG 109419  
QY 421 GCGGTTGAGCGATCCGCTCGATACACGAGCGGTTCGCGGCTCGCGGCGGATC 480  
Db 109420 GCGGTTGAGCGATCCGCTCGATACACGAGCGGTTCGCGGCTCGCGGCGGATC 109479  
QY 481 GGTACCGCGGACCGCTCGCGCAGCGCGCGCGGATCC 520  
Db 109480 GGTACCGCGGACCGCTCGCGCAGCGCGCGCGGATCC 109519

## RESULT 2

AX196102/c AX196102 579 bp DNA linear PAT 28-AUG-2001

DEFINITION Sequence 174 from Patent WO0151639.

ACCESSION AX196102

VERSION AX196102.1 GI:15386334

## SOURCE

ORGANISM Micromonospora carbonacea

Micromonospora carbonacea

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Micromonosporineae; Micromonosporaceae; Micromonospora.

## REFERENCE

1 Hosted, T.J., Horan, A.C. and Wang, T.X.  
Evernimycin biosynthetic genes  
Patent: WO 0151639-A 174 19-JUL-2001;  
Schering Corporation (US)

## FEATURES

source

1. 109519  
/organism="Micromonospora carbonacea"

/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

## CDS

1..579

/note="unassigned protein product; evrMR2"

/codon\_start=1

/transl\_table=11

/protein\_id="CAC60077.1"

/db\_xref="GI:15386335"

/db\_xref="RENTREMBL:CAC60077"

/translation="MWSPGDEIEURPTVEIARTAVLHVILARCFGMPAQAAMSWLLN  
SHLDVLTPEPWFVGMGRDHSFVLAHDALESLAMVGLAKOLDTLPVDERLVER  
LPHIAGSEFCQNRSLIAPQHPADAPALDLHYCLDWAYLELDRGRSPFGPLVDAN  
AIGQRWALEWAVVIRGPIHEEPGWEVDLST"

## ORIGIN

Query Match 41.7%; Score 217; DB 6; Length 579;

Best Local Similarity 100.0%; Pred. No. 9.6e-23;

Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGACAGCAACGACCGGTGGTTCGCCCGCGCCATCAGAACTGCCTCCGGCG 60  
Db 217 GGTGACAGCAACGACCGGTGGTTCGCCCGCGCCATCAGAACTGCCTCCGGCG 158  
QY 61 GGTGACAGGTGACAGGTGGAGTTGAGCAGCCAGTCTATCGCGCTTCGGCGGCA 120  
Db 157 GGTGACAGGTGACAGGTGGAGTTGAGCAGCCAGTCTATCGCGCTTCGGCGGCA 98  
QY 121 TGCCGAAACACCGCGCCAGGATCAGTGCAGCAGCGGATCGCGCTCGATCTCCAGCG 180  
Db 97 TGCCGAAACACCGCGCCAGGATCAGTGCAGCAGCGGATCGCGCTCGATCTCCAGCG 38  
QY 181 TCGGCGCAGTCTCGTTCGCCCGCGTCCACACAC 217  
Db 37 TCGGCGCAGTCTCGTTCGCCCGCGTCCACAC 1

## RESULT 3

E02937

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NEW GENE MANIFESTING MACROLIDE ANTIBIOTIC RESISTANCE

Patent JP 199139284-A 1 13 JUN 1991

TOYO JOZO CO LTD

OS Micromonospora glisoelebia

PN JP 1991139284-A/1

PD 13-JUN-1991

PF 28-MAY-1990 JP 1990137997

PR 11-JUL-1989 JP 89P 178490

PI INOUE MASAHARU, SUZUKI YASUSHI, MOROHOSHI TOSHIRO, MUTO NAKI,

PC HORINOCHI SUEJI, BEPPU TERUHIKO

PC C12N15/65, C12N1/21, C12N15/76, (C12N15/65, C12N1/29), (C12N1/21,

PC C12N1/465)

CC (C12N15/76, C12N1/465);

CC strandedness: Double;

CC topology: Linear;

CC hypothetical: No;

CC anti-sense: No;

CC \*source: strain=All1725(FERM BP-705);

CC key Location/Qualifiers

FT misc\_feature >1..<1954

FT Location/Qualifiers

FT /gene='Mycaminacin resistant gene'.

FT 1..1954

source

	/organism="unidentified"						
	/mol_type="genomic DNA"						
	/db_xref="taxon:32644"						
ORIGIN							
Query Match	16.6%;	Score 86.2;	DB 6;	Length 1954;			
Best Local Similarity	81.3%;	Pred. No. 0.0012;	Mismatches 0;	Gaps 0;			
Matches 100;	Conservative	0;	Mismatches 23;	Indels 0;	Gaps 0;		
QY	1	GGTGAGCAGCAACGACCCTGGTGCCTCCCGCCGCCCATCACGAATGCCACTTCGCGCG 60					
Dd	1832	GGTGAGCAGCAACGAAAGCGTGGTGCCTCCCGCGCGGATGATGAAGTCCACTTCGCGCG 1891					
QY	61	GGGTGACAGGTGACGACCGAGTTGGAGCAGCAGCTATCCCGCCTGCCTGCCCGCA 120					
Dd	1892	GGGTGACAGGTGACGACCGAGTTGGAGCAGCAGCTATCCCGCCTGCCTGCCCGCA 1951					
QY	121	TGC 123					
Dd	1952	TGC 1954					
RESULT 4							
AC139773/c	55351 bp DNA linear HTG 13-FEB-2003						
LOCUS	Homo sapiens chromosome 19 clone LLNLF-200C8, WORKING DRAFT						
DEFINITION	SEQUENCE, 22 unordered pieces.						
ACCESSION	AC139773.1	GI:28372800					
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.						
KEYWORDS	Homo sapiens (human)						
SOURCE	Homo sapiens						
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
REFERENCE	1 (bases 1 to 65351); DOE Joint Genome Institute.						
AUTHORS	Unpublished Sequencing of Human Chromosome 19						
JOURNAL	DOE Joint Genome Institute.						
REFERENCE	2 (bases 1 to 65351); Direct Submission.						
AUTHORS	Submitted (13-FEB-2003) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA						
TITLE	-----Genome Center						
JOURNAL	Center: Joint Genome Institute						
COMMENT	Center Code: JGI						
	Web site: http://www.jgi.doe.gov						
	-----						
Project Information							
Center Project Name:	19502, F24136						
Center clone name:	LLNLF_200C8						
Summary Statistics							
Consensus quality:	44919 bases at least Q40						
Consensus quality:	48008 bases at least Q30						
Consensus quality:	51090 bases at least Q20						
Estimated insert size:	37980; agarose-fp estimation						
Estimated insert size:	63251; sum-of-contigs estimation						
Quality coverage:	37.55 in Q20 bases; agarose-fp estimation						
Quality coverage:	22.55 in Q20 bases; sum-of-contigs estimation.						
* NOTE:	This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.						
*	1	1361:	contig of 1361 bp in length				
*	1362	1461:	Gap of unknown length				
*	1462	2597:	contig of 1136 bp in length				
*	2598	2697:	Gap of unknown length				
*	2698	3861:	contig of 1164 bp in length				
FEATURES							
Location/Qualifiers							
1..65351							
/organism="Homo sapiens"							
/mol_type="genomic DNA"							
/db_xref="taxon:9606"							
/chromosome="19"							
/clone="LLNLF-200C8"							
/clone_lib="Lawrence Livermore human cosmid library LLNLF"							
ORIGIN							
Query Match	12.4%;	Score 64.6;	DB 2;	Length 65351;			
Best Local Similarity	45.3%;	Pred. No. 0.751;	Mismatches 223;	Conservative 0;	Mismatches 269;	Indels 0;	
Matches 223;	Conservative	0;	Mismatches 269;	Indels 0;	Gaps 0;		
QY	25	CGCCCCCGCGCGCCATCACGAATCGCTCCCTCCCGCGCGGCTGACACAGGTGCACGAGTGGG 84					
Dd	2296	CGCCCCCGCGCGCGCCATCACGAATCGCTCCCTCCCGCGCGGCTGACACAGGTGCACGAGTGGG 223					
QY	85	AGTTGAGCAGCAGCTATCCCGCCTGCCTCCCGCGCGGCTGACACAGGTGCACGAGTGGG 144					
Dd	2236	CGCCCCCGCGCGCGCCATCACGAATCGCTCCCTCCCGCGCGGCTGACACAGGTGCACGAGTGGG 2177					
QY	145	CGTGCAGCAGCGCGATGCGCGCTCGATCTCGACCGTCCGCGCGGCTCGAGTCTGCTCGC 204					
Dd	2176	CGCCCCCGCGCGCGCCATCACGAATCGCTCCCTCCCGCGCGGCTGACACAGGTGCACGAGTGGG 2117					
QY	205	CGCGCTCCACACACAGGGAATACTGGTCTGGTGGCAGCGGAGCCCACAGCGCGGACACT 264					
Dd	2116	CGCGCGCGCGCGCGCCATCACGAATCGCTCCCTCCCGCGCGGCTGACACAGGTGCACGAGTGGG 205					
QY	265	CGTTCAGGCTGGCGTCAAGGACTCTTCGCGCGGCTCCGAGCGGAACAGCGCTCAGA 324					
Dd	2056	CGCCCCCGCGCGCGCCATCACGAATCGCTCCCTCCCGCGCGGCTGACACAGGTGCACGAGTGGG 1997					
QY	325	TCCTCTCAGTCGCAATCGGCTCAGTCCCGGCTCGTCCCTTTGGCTGGAGATAGCGGT 384					





Db 37855 CGTTGGCGGTCAACGGACCCGCTCTCCAGCACACCAGCGGCCG 37806

RESULT 8  
SCO939118/c

LOCUS SCO939118 303550 bp DNA linear BCT 11-FEB-2003

DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 15/29.

ACCESSION ALJ93118 AL034355 AL049826 AL079308 AL098823 AL118514 AL118515  
ALJ56612 ALJ57152 ALJ59217 ALJ59394 ALJ59625 ALJ64582  
ALJ93118.1 GI:124413861

VERSION .

KEYWORDS Streptomyces coelicolor A3(2)

SOURCE Streptomyces coelicolor A3(2)

ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
AUTHORS Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,  
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,  
Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,  
Cranin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,  
Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,  
Rabinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,  
Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S., B.G.,  
Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrall,B., Parkhill,J.  
and Hopwood,D.A.

TITLE Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)

JOURNAL Nature 417 (6895), 141-147 (2002)

MEDLINE 21996410

PUBMED 12000953

REFERENCE 2 (bases 1 to 303550)  
Bentley,S.D.  
Direct Submission  
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces  
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
On or before Oct 26, 2002 this sequence version replaced  
GI:20520812, GI:4808374, GI:5102782, GI:5457267, GI:5918466,  
GI:5918500, GI:8052426, GI:8247645, GI:10178365, GI:13122127,  
GI:13122160, GI:15021246.

COMMENT Location/Qualifiers  
1. 303550  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="Genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
424..1296  
/gene="SCO3843"  
/note="synonym: SCH69.13"  
424..1296  
/gene="SCO3843"  
/note="SCH69.13, unknown, len: 290 aa; similar to the  
N-terminal region of Tr:P71590 (EMBL:Z80233) hypothetical  
protein from Mycobacterium tuberculosis (527 aa) fasta  
scores; opt: 268, z-score: 221.4, E(): 5.3e-05, (28.1%  
identity in 217 aa overlap)."  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAB45210.1"  
/db\_xref="GI:5102795"  
/db\_xref="GOA:O9XA21"  
/db\_xref="SwProt:P09XA21"  
/translation="MGVLKKEPQRLGLVNGTFAKVPSVQPVEIAGNLQRECDNNNA  
TIWNRDITVPNDIFSVLSFTDFERLSPYSGQLGDELACGMVRDYAKQRITYFMGPVKI  
HLEKADLDLTGLKYRSTRILASSTSSPGAPAAAPAPPAPPPAGGGGYGPPPAALAI  
GAPMPAPPPGPRGGYGYPQAPAGGYFPQAPAGGTPAAPAPGGRTRWIEINGTRHQISRAII  
MGRSTEADVRIIDFGVSRRRHCIEIRTGTFPTSIQDLGSTNGIVVDQGHTTRATLRDGSGT  
VGVTWTYRQAEG"  
1307..1825  
/gene="SCO3844"  
/note="synonym: SCH69.14"  
1307..1825

FEATURES  
source

gene

CDS

gene

CDS



```

/gene="SCO3844"
/notes="SCH69.14, putative secreted protein, len: 172 aa;
similar to TR:P71589 (EMBL:Z80233) hypothetical protein
from Mycobacterium tuberculosis (155 aa) fasta scores;
opt: 389, z-score: 442.8, E(): 2.5e-17, (46.2% identity in
171 aa overlap). Contains possible N-terminal region
signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAB45211.1"
/db_xref="GI:5102796"
/db_xref="SPTREMBL:Q9XA20"
/translation="MSLITLWMLGFLAVLWLVIVAVIQRSDLGTEVVGARR
DARPQATRGQAPFPKQSCGGRGRGATKLVTEGTLITVALQGTITLG
RAHSTPILDDDDYASSRHARIYPQNGQWIVDLGSLTYDLRARLTTPTPGPGAP
IRIGKTVIELRK"
1967. 3514
/gene="SCO3845"
/notes="synonym: SCH69.15"
1967. 3514
/gene="SCO3845"
/notes="SCO3845"
similar to many of undefined function egs. TR:P71588
(EMBL:Z80233) hypothetical protein from Mycobacterium
tuberculosis (514 aa) fasta scores; opt: 989, z-score:
916.0, E(): 0. (42.9% identity in 536 aa overlap) and
TR:Q50188 (EMBL:Z70722) putative phosphoprotein phosphatase
from Mycobacterium leprae (509 aa) fasta scores; opt: 966,
z-score: 895.0, E(): 0. (40.0% identity in 530 aa
overlap). Contains 2 Pfam matches to entry PF00481 PP2C,
protein phosphatase 2C. could have a role in signalling."
/codon_start=1
/transl_table=11
/product="putative protein phosphatase"
/protein_id="CAB45212.1"
/db_xref="GI:5102797"
/db_xref="GOA:Q9XA19"
/db_xref="SPTREMBL:Q9XA19"
/translation="MSLSLRPAAGSHKGMIREGNSGYAGPRLAIADGGQQAAGE
VASSEVITVALDDVPGSDVLTSGLHAVORANDOLRWVEEDPALEGMTTLTALL
WTQRLGVHVGDSRAVLRDGLVLTQDHTWVRLVDEGRITEREAGTHPORSILM
RALGSGDHFPLSIREVRAGDRYLICSDGLSGVSVQTHMEDILASQGSQETVQELI
QLAURGQGVNITVIVADVLDDLTGLAGLSDTVPVGAVAENQAQMGDGLMQLTP
AGRAAGLRPGQGGGEGFPGSGDVTGFIAGDDFDYDFDVFPRKSRWLKRS
LYTALAALVGGTGGGWRWTQTYVTGTTNDHLALYGISQDLAWLSLKVQKHDE
IELKLPYQOKLVEATIPEDGLNDAKIEBELAQASACKQAARATAEKNAKTG
EGEAGTGTGTPASFTSKASPSPNPGSPAPESSESPSTTATPPGSGPTLSBEEQKV
VSLCGKO"
1979. 2368
/gene="SCO3845"
/notes="Pfam match to entry PF00481 PP2C, Protein
phosphatase 2C, score 32.60, E-value 2.8e-08."
2474. 2656
/gene="SCO3845"
/notes="Pfam match to entry PF00481 PP2C, Protein
phosphatase 2C, score 9.50, E-value 0.11."
3542. 4981
/gene="SCO3846"
/notes="synonym: SCH69.16"
3542. 4981
/gene="SCO3846"
/notes="SCH69.16, probable FtsW/RodA/Spove family cell
cycle protein, len: 479aa; similar to many egs.
Sw:FTSW MYCUT probable cell division protein FtsW from
Mycobacterium tuberculosis (469 aa) fasta scores; opt:
1367, z-score: 1568.6, E(): 0. (48.8% identity in 451 aa
overlap) and Sw:SPSE BACSU stage V sporulation protein E
from Bacillus subtilis (366 aa) fasta scores; opt: 496,
z-score: 572.9, E(): 1.4e-24, (31.7% identity in 391 aa
overlap). Also similar to SC6G9.31 from Streptomyces
coelicolor (446 aa) fasta scores; opt: 1132, z-score:
1182.0, E(): 0. (58.6% identity in 444 aa overlap).

```

```

Contains Pfam match to entry PF01098 FTSW RODA SPOVE, Cell
cycle protein. Also contains possible membrane spanning
hydrophobic regions."
/codon_start=1
/transl_table=11
/product="putative FtsW/RodA/Spove family cell cycle
protein"
/protein_id="CAB45213.1"
/db_xref="GI:5102798"
/db_xref="SPTREMBL:Q9XA18"
/translation="MSSTNFTSTHTSTIGAIPSRRTLALVFAVLIPVFAYAN
VGLAINDVPAGLLSYGFLGGLAGVHVVRKFPAPYADPLLLPLATLNGILVIAIW
RLDQSDRLQSSKLPVEAAPROMLYSALGVAVLRFDRHVQLQRTVYISMAQALV
LLLLPLGLGHDFNGAKIWKIPGLGTLOQBEPAKIVLAVFAGVLMVKRDLALAS
RFLMGLYLPGRDGLGPIIVMTISLILVAFSTDLGTSLLFGFMVIMLVATERTSWI
VFLGMSAVAGVAVSFESHVQORVOAWLDPKHVELSRQGVFGHTESQMQLWAPGS
GDTLGSWGQGSNDLIGFAANSDFILATFGBELGLAGLMALLLILVILVERVTRALA
ARDPFGKLLATLGLSAGFALQVVFAGGVNGLPLTGMTMPFLAYGGSSVIANWALIGI
LLRISDTRRFPAPAPANPDAMTQVRP"
3788. 4921
/gene="SCO3846"
/notes="Pfam match to entry PF01098 FTSW RODA SPOVE, Cell
cycle protein, score 253.10, E-value 3.7e-72."
4978. 6450
/gene="SCO3847"
/notes="synonym: SCH69.17"
4978. 6450
/gene="SCO3847"
/notes="SCH69.17, probable secreted penicillin-binding
protein, len: 490 aa; similar to many e.g. TR:P71586
(EMBL:Z80233) putative penicillin-binding protein from
Mycobacterium tuberculosis (491 aa) fasta scores; opt:
112.1 z-score: 1262.1, E(): 0. (39.3% identity in 499 aa
overlap) and SW:P8PC BACSU penicillin-binding protein 3
from Bacillus subtilis (668 aa) fasta scores; opt: 447,
z-score: 502.5, E(): 1.2e-20, (29.0% identity in 386 aa
overlap). Also similar to SC6G9.32 from Streptomyces
coelicolor (485 aa) fasta scores; opt: 1493, z-score:
1531.4, E(): 0. (49.4% identity in 488 aa overlap).
Contains Pfam match to entry PF00905 transpeptidase,
penicillin binding protein transpeptidase domain. Also
contains possible N-terminal signal sequence."
/codon_start=1
/transl_table=11
/product="putative secreted penicillin-binding protein"
/protein_id="CAB45214.1"
/db_xref="GI:5102799"
/db_xref="GOA:Q9XA17"
/db_xref="SPTREMBL:Q9XA17"
/translation="VNKPLRRIALFGLLVLLIRDNWLVQVQADELKHEDHNRVVA
IRYASPRGDIIVDGKAITGHATTEGDFKXKTKYKDGAMWAPVTVGSVAFGNTQLES
IDGLITGNDDRLDFRNTLDMITGKRGGVNVTTLNGAQAADGLKQKAGKAVV
AIEPSTGKILSMASYSYDPTVIAGSNEAAGWANKLOKKNPDPMMLNRLREVTPP
GSTFKVYTAALAEHGKYASADEKTSPLPWTMPGTDTLKNEGNI PCENATLREALR
VSCNTVFGKLYDLGLNKMLETAKFTGTEQFVFRSSASVFSMDMDSQVALSSIG
QFNATATPLQMAWTSALANNKTLMKPYMVBDELQGLDTEKTPDEENSEPLSADNA
QILQSMETVVEDGTGNAKIDGVKVGKTAQHGVDNSENPNYAFWIFSYAKADDGSS
PVAVAVIEDENAVRDDISGGSLAAPLAKWMEAVIDSKK"
5395. 6417

```

```

misc_feature
Query Match 11.4%; Score 59.4; DB 1; Length 303550;
Best Local Similarity 46.7%; Pred. No. 3;
Matches 221; Conservative 0; Mismatches 251; Indels 1; Gaps 1;

Qy 48 TGCCTACGCGCGGTGACAGGTGACAGGTGAGTTCAGCAGCGAGTTCAGCAGCGCTCATCGCC 107
Db 289969 TGCCTACGCGCGGTGACAGGTGACAGGTGAGTTCAGCAGCGAGTTCAGCAGCGCTCATCGCC 289910
Qy 108 GCCTGCGCGCGGTGACAGGTGACAGGTGAGTTCAGCAGCGAGTTCAGCAGCGCTCATCGCC 167
Db 289909 GCCTGCGCGCGGTGACAGGTGACAGGTGAGTTCAGCAGCGAGTTCAGCAGCGCTCATCGCC 289850
Qy 168 TCGATCTGACGGTTCGCGCGGTGACAGGTGAGTTCAGCAGCGAGTTCAGCAGCGCTCATCGCC 227

```



```
18633 20007: contig of 1375 bp in length
* 20008 20107: gap of unknown length
* 20108 21365: contig of 1258 bp in length
* 21366 21465: gap of unknown length
* 21466 22492: contig of 1027 bp in length
* 22493 23648: contig of 1056 bp in length
* 23649 23748: gap of unknown length
* 23749 25101: contig of 1353 bp in length
* 25102 25201: gap of unknown length
* 25203 26496: contig of 1295 bp in length
* 26497 28596: gap of unknown length
* 28597 28576: contig of 1980 bp in length
* 28577 28676: gap of unknown length
* 28677 30152: contig of 1476 bp in length
* 30153 30252: gap of unknown length
* 30253 31814: contig of 1562 bp in length
* 31815 31914: gap of unknown length
* 31915 33202: contig of 1388 bp in length
* 33203 33402: gap of unknown length
* 33403 34583: contig of 1181 bp in length
* 34584 34683: gap of unknown length
* 34684 36574: contig of 1891 bp in length
* 36575 36674: gap of unknown length
* 36676 38144: contig of 1470 bp in length
* 38145 38244: gap of unknown length
* 38245 40273: contig of 2029 bp in length
* 40274 40373: gap of unknown length
* 40374 41883: contig of 1510 bp in length
* 41884 41983: gap of unknown length
* 41984 44125: contig of 2142 bp in length
* 44126 44225: gap of unknown length
* 44226 45503: contig of 2278 bp in length
* 45504 46603: gap of unknown length
* 46604 48746: contig of 2143 bp in length
* 48747 48846: gap of unknown length
* 48847 51418: contig of 2572 bp in length
* 51419 51518: gap of unknown length
* 51519 54977: contig of 3459 bp in length
* 54978 55077: gap of unknown length
* 55078 57345: contig of 2268 bp in length
* 57346 57445: gap of unknown length
* 57446 59735: contig of 2290 bp in length
* 59736 61781: gap of unknown length
* 61782 61881: contig of 1946 bp in length
* 61882 63845: contig of 1964 bp in length
* 63846 63945: gap of unknown length
* 63946 65199: contig of 1254 bp in length
* 65200 65299: gap of unknown length
* 65300 66555: contig of 1356 bp in length
* 66556 66755: gap of unknown length
* 66756 69589: contig of 2834 bp in length
* 69590 69689: gap of unknown length
* 69690 72230: contig of 2541 bp in length
* 72231 72330: gap of unknown length
* 72331 74676: contig of 2346 bp in length
* 74677 74776: gap of unknown length
* 74777 76884: contig of 2108 bp in length
* 76885 76984: gap of unknown length
* 76985 79581: contig of 2597 bp in length
* 79582 79681: gap of unknown length
* 79683 83305: contig of 3624 bp in length
* 83306 83405: gap of unknown length
* 83406 85985: contig of 2480 bp in length
* 85986 85986: gap of unknown length
* 85987 89565: contig of 3580 bp in length
* 89566 93622: gap of unknown length
* 93623 93627: gap of unknown length
* 93628 96375: contig of 3113 bp in length
* 96376 99096: contig of 2721 bp in length

Query Match 11.1%; Score 57.6; DB 2; Length 129506;
Best Local Similarity 42.9%; Pred. No. 6.8;
Matches 213; Conservative 0; Mismatches 284; Indels 0; Gaps 0;
QY 24 TCGCCCGCGCGCCATCAGAACTGCGACTCGGGGGGGTACAGGTGACAGGTGG 83
DB 16460 TCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 16519
QY 84 GAGTTGAGCAGCAGCTCATCGCGCGCTCGCGCGCATCCGAAACACCGGCGCCAGATC 143
DB 16520 CCNCCNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16579
QY 144 AGGTGAGCAGCGCGATGCGCGCTCGATCTCGAGGTGCGCGCGCGAGTCTGATCTGTCG 203
DB 16580 GGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16639
QY 204 CCGGCTCCACACACAGGGGAACTGGCTCGGTGCGAGCGCGCGCGCGCGCGCGCGCG 263
DB 16640 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16699
QY 264 TCGTCCAGGCTGCGGTACGAGCTTTCGCGGGTGGGAGCGGAAACGCGCGCGTCTAG 323
DB 16700 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16759
QY 324 ATCCCTGTGATCGCATCGGCTCAGTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 383
DB 16760 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16819
QY 384 TTCACGACGAGCGGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 443
DB 16820 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16879
QY 444 GACACGCGGTGCGGGGTGCGGGCGGTGCGGAGATCGTACGCGCGCGCGCGCGCGCG 503
DB 16880 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16939
QY 504 CAGCGCGCGCGCGGATCC 520
DB 16940 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16956

RESULT 10
SCO939119/c
LOCUS SCO939119 299050 bp DNA linear BCT 11-FEB-2003
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 16/29.
ACCESSION AL939119 AL160331 AL353816 AL356612 AL357432 AL357524 AL358692
AL358818 AL359988 AL391406 AL391541 AL392149 AL392150 AL392178
AL645882
VERSION AL939119.1 GI:24427855
KEYWORDS Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomyces; Streptomyces.
REFERENCE 1
AUTHORS Bentley,S.D., Chater,K.F., Cerdano-Tarraga,A.M., Challis,G.L.,
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
Cronin,A., Frazer,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,
Rabinowitz,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrall,B.G.,
Parkhill,J. and Hopwood,D.A.
Complete genome sequence of the model actinomycete Streptomyces
```

coelicolin A3(2)  
Nature 417. (6885), 141-147 (2002)  
21996410  
12000953  
2. (bases 1 to 299050)  
Bentley, S.D.  
Direct Submission  
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces  
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
On or before Oct 30, 2002 this sequence version replaced  
gi:7242748, gi:7636006, gi:8248766, gi:8249971, gi:8388712,  
gi:8439477, gi:8894718, gi:9857143, gi:9886716, gi:10129746,  
gi:11595528.

FEATURES  
source  
1. .299050  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
200..1727  
/genes="rRNA"  
/genes="rRNA"  
/notes="rRNA, 16S ribosomal RNA. Coordinates based on  
comparison with rRNA from Streptomyces coelicolor  
(EMBL:AL109848)."  
2004..5126  
/genes="rRNA"  
2004..5126  
/genes="rRNA"  
/notes="rRNA, 23S ribosomal RNA. Coordinates based on  
comparison with rRNA from Streptomyces coelicolor  
(EMBL:AL109848)."  
5228..5348  
/genes="rRNA"  
/genes="rRNA"  
/notes="rRNA, 5S ribosomal RNA. Coordinates based on  
comparison with rRNA from Streptomyces coelicolor  
(EMBL:AL109848)."  
complement(5412..6080)  
/gene="SCO4123"  
/note="synonym: SCD72A.09c"  
complement(5412..6080)  
/genes="SCO4123"  
/notes="SCD72A.09c, possible two-component system response  
regulator, len: 222aa; similar to many eg. TR:086632  
(EMBL:AL031155) possible two-component regulator from  
Streptomyces coelicolor (222 aa) fasta scores; opt: 973,  
z-score: 1128.5, E(): 0, 69.1% identity in 220 aa overlap.  
Contains Pfam match to entry PF00072 response\_reg,  
Response regulator receiver domain, Pfam match to entry  
PF00196 GerE, Bacterial regulatory proteins, luxR family  
and Prosite match to PS00622 Bacterial regulatory  
proteins, luxR family signature."  
/codon\_start=1  
/transl\_table=11  
/product="putative two-component system response  
regulator"  
/protein\_id="CAB92376.1"  
/db\_xref="GI:8052435"  
/db\_xref="GOA:Q9KYB7"  
/db\_xref="SPTRMBL:Q9KYB7"  
/translation="NTVLVADQPLQRFGEFMLESDQMTVGEANGAARMAA  
ELHPDVLMVRMPGLDGEATRIIVAGGRSKVLITFTDDEYASIGSRAGAGGFL  
VKDALPEELSGSRATVADAVAPSLTRLLDAYVHLPAAAGTRDHTPGSPDRLSP  
LTERETELFVWGQGWNTIEAERLHAESTVTKVHGRILAKTGRDRVQAVILAYET  
KLVG"  
complement(5415..5609)  
/gene="SCO4123"  
/note="Pfam match to entry PF00196 GerE, Bacterial  
regulatory proteins, luxR family, score 61.00, E-value

2.6e-14"  
complement(5475..5558)  
/gene="SCO4123"  
/note="PS00622 Bacterial regulatory proteins, luxR family  
signature"  
complement(5736..5077)  
/genes="SCO4123"  
/note="Pfam match to entry PF00072 response\_reg, Response  
regulator receiver domain, score 126.30, E-value 5.7e-34"  
complement(6077..7447)  
/gene="SCO4123"  
/note="synonym: SCD72A.10c"  
complement(6077..7447)  
/genes="SCO4123"  
/notes="SCD72A.10c, possible two-component sensor kinase,  
len: 456aa; similar to many eg. TR:086631 (EMBL:AL031155)  
putative two-component sensor kinase from Streptomyces  
coelicolor (429 aa) fasta scores; opt: 929, z-score:  
995.1, E(): 0, 42.7% identity in 422 aa overlap. Contains  
Pfam match to entry PF00512 signal, Histidine kinase and  
possible membrane-spanning hydrophobic regions. Contains  
possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/transl\_table=11  
/product="putative two-component sensor kinase"  
/protein\_id="CAB92377.1"  
/db\_xref="GI:8052436"  
/db\_xref="GOA:Q9KYB6"  
/db\_xref="SPTRMBL:Q9KYB6"  
/translation="MEATVIFGYAPHLYIGPTDARPARTPDHGMPEVKVPEDRA  
AEPTEYQVRNRIIRAFDRHFLLDLTGFWLAALVDGDMRSVAHDGPGVPV  
GLVLVLSALSVFVVRSHPRGALLITPAALVSAWSGAALQALLVLLHIALR  
LPRLNLMVAVVAAPTCSVYSGSAGSNQQLSQAIVAGALIGITVTRRGY  
TEALDRALLETQRDOQVSLAAARARARAHMDIIGHNLVITLADGGVAAAK  
SPDRAQALDAIGTTSQALGELRLDLVLRDDQDGTGPELAPQALTDRLDLD  
GVRAGLPVTTVKGEPLPEGRQLTVRVIQEALNTLKHAGPDATAHLELYSGSG  
AALTVTDGTGRTAPLSPGSGRGLPGMRKTSLSGLTLEAGRPAEQGMRVHLLH  
PEESFQ"  
complement(6095..6670)  
/gene="SCO4124"  
/note="Pfam match to entry PF00512 signal, Histidine  
kinase, score -41.40, E-value 0.0087"  
7473..8006  
/gene="SCO4125"  
/note="synonym: SCD72A.11"  
7473..8006  
/gene="SCO4125"  
/note="SCD72A.11, possible acetyltransferase, len: 177 aa;  
weakly similar to many eg. TR:AD52985 (EMBL:AF140221)  
aminoglycoside 6'-N-acetyltransferase from Xanthomonas  
maltophilia (153 aa) fasta scores; opt: 128, z-score:  
164.9, E(): 0.1, 28.7% identity in 115 aa overlap.  
Contains Pfam match to entry PF00583 Acetyltransf,  
Acetyltransferase (GNAT) family."  
/codon\_start=1  
/transl\_table=11  
/product="putative acetyltransferase"  
/protein\_id="CAB92378.1"  
/db\_xref="GI:8052437"  
/db\_xref="GOA:Q9KYB5"  
/db\_xref="SPTRMBL:Q9KYB5"  
/translation="MIRSHADEWPAVKELRLRALRDPVADLAYLTFEEAARPDYSY  
WQEARAGSAGCAQOVAVPGGLWLTGLTVLVEAGTTDWAGFVRRQGHVGV  
FVREPWGSLTKALEFDAGLENGWANGVERMLIVHEGNTAQGAYKAGFAPSGRTV  
PLEGPGDHLLEVYLER"  
7476..7934  
/gene="SCO4125"  
/note="Pfam match to entry PF00583 Acetyltransf,  
Acetyltransferase (GNAT) family, score 34.50, E-value  
2.4e-06"  
complement(8054..9541)  
/gene="SCO4126"  
/note="synonym: SCD72A.12c"

misc\_feature  
complement(8054..9541)  
/gene="SCO4126"  
/note="synonym: SCD72A.12c"



```
misc_feature 87808..96100
/notes="assembly_fragment:01700
fragment_chain:2"
misc_feature 96201..98287
/notes="assembly_fragment:00028
fragment_chain:2"
misc_feature 98388..102420
/notes="assembly_fragment:00012
fragment_chain:2"
misc_feature 102521..1168071
/notes="assembly_fragment:00976
fragment_chain:3"
misc_feature 168172..179602
/notes="assembly_fragment:01864
fragment_chain:3"
misc_feature 179703..181850
/notes="assembly_fragment:01446"

ORIGIN
Query Match 10.8; Score 56.2; DB 2; Length 181850;
Best Local Similarity 48.78; Pred No 10;
Matches 243; Conservative 0; Mismatches 248; Indels 8; Gaps 3;

Qy 14 CGACCGGTGTGCGCCCGCGCGCCGACGCACTGCGCGGCGGTCGCGGAGTGCACGAGTC 3
Db 4306 CCACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4365
Qy 74 GACCAAGTGGAGTTGAGCAGCAGCAGCTCATCGCGGCTGCGCGGCGATGCCGAAACACCG 133
Db 4366 GCGCGGGGGGGGGCGCGCGCTCGGGGGGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 4423
Qy 134 GGCCAGGATCAGCTGACGACGCGGATGCGCGGCTCGATCTGACGCTGCGCGCGCAGTC 193
Db 4424 GCCAGGGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4483
Qy 194 GATCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 253
Db 4484 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4543
Qy 254 CCGCGGACAGTGTGCTTCAGGCTGCGGTTCACGACGCTCTTCGCGGCTGCGGAGCGGAAACGCG 313
Db 4544 GCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4597
Qy 314 CACGCGTCAGATCCTGTGTCAGTCGATCGGTCAGTCGCGGTGTCGCGGTCGCGGTCGCG 373
Db 4598 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4657
Qy 374 AGGATAGCGGTTCACGACGAGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
Db 4658 GCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4717
Qy 434 TCCGCTCGATGACGAGCGGCTGCGGCGGTGCGGCGGTGCGGCGAGATCCGTACCGCGCGCG 493
Db 4718 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4777
Qy 494 CCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 512
Db 4778 GCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4796

RESULT 12
AC079420
LOCUS
DEFINITION Mus musculus chromosome 15 clone RP23-104L12, WORKING DRAFT
SEQUENCE, 61 unordered pieces.
ACCESSION AC079420.1 GI:9958032
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
MUS musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 239130)
```

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

DOE Joint Genome Institute.  
Sequencing of Mouse  
Unpublished  
2 (bases 1 to 239130)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Project Name: 0  
Center clone name: RPCI-23\_104L12  
-----

## Summary Statistics

Consensus quality: 154414 bases at least Q40  
Consensus quality: 183510 bases at least Q30  
Consensus quality: 194836 bases at least Q20  
Estimated insert size: 222300; agarose-fp estimation  
Estimated insert size: 233130; sum-of-contigs estimation  
Quality coverage: 3.27 in Q20 bases; agarose-fp estimation  
Quality coverage: 3.12 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 61 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1030: contig of 1030 bp in length  
1031 1130: gap of unknown length  
1131 2520: contig of 1390 bp in length  
2521 2620: gap of unknown length  
2621 3633: contig of 1013 bp in length  
3634 3734: gap of unknown length  
3735 5381: contig of 1648 bp in length  
5382 5482: gap of unknown length  
5483 6792: contig of 1311 bp in length  
6793 8478: contig of 1586 bp in length  
8479 8578: gap of unknown length  
8579 9634: contig of 1056 bp in length  
9635 9734: gap of unknown length  
9735 10791: contig of 1057 bp in length  
10792 10891: gap of unknown length  
10892 12258: contig of 1367 bp in length  
12259 12358: gap of unknown length  
12359 13771: contig of 1413 bp in length  
13772 13871: gap of unknown length  
13872 14998: contig of 1127 bp in length  
14999 15098: gap of unknown length  
15099 16261: contig of 1163 bp in length  
16262 16361: gap of unknown length  
16362 17426: contig of 1065 bp in length  
17427 17526: gap of unknown length  
17527 18541: contig of 1015 bp in length  
18542 18641: gap of unknown length  
18642 20200: contig of 1559 bp in length  
20201 20300: gap of unknown length  
20301 21423: contig of 1123 bp in length  
21424 21523: gap of unknown length  
21524 22553: contig of 1030 bp in length  
22554 22853: gap of unknown length  
22854 23664: contig of 1010 bp in length  
23665 23763: gap of unknown length  
23764 24815: contig of 1052 bp in length  
24816 24915: gap of unknown length  
24916 26356: contig of 1441 bp in length  
26357 26456: gap of unknown length



AUTHORS	Hattori, M., Iahii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiyama, A., Ikada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (04-DEC-2000) Maashira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Japan (E-mail: hattori@ruc.riken.go.jp, URL: <a href="http://hgp.gsc.riken.go.jp/">http://hgp.gsc.riken.go.jp/</a> , Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT	On Jun 16, 2003 this sequence version replaced gi:21623976.
FEATURES	Location/Qualifiers
source	1. .191589

ORIGIN	Query Match	10.8%;	Score 56;	DB 9;	Length 191589;
	Best Local Similarity	48.6%;	Pred. No.11;		
	Matches 245;	Conservative 0;	Mismatches 250;	Indels 9;	Gaps 3;
QY	3	TCGACGACGAA	CGACCGGTGTTGCCCCCGCGGCCCATCACGAACTGCACCTCCGGCGGG	62	
Db	49398	TTCTGCACCGGGGCGAGCGAGCCCTCGCGCTCGGCTCTCTCTGAGTGCGGTAGC	49339		
QY	63	GTGACCAAGTTCGACCAAGGTGGAGTTGACAGCCAGTTCATCGCGCCTTCGCGCGGATG	122		
Db	49338	GGCAGCGCGGCCCGCGCGCCGGCTGGGGGACGACGCGGACCGCGCAGGACCTCGAG	49279		
QY	123	CCGAAACACGGGCGCCAGGATCAGTCGACACGGCGATGCGCGCTTCGATCTTCGACGGTC	182		
Db	49278	TCCAAGCAGCTGGGGCGG----	49224		
QY	183	GGCGCGAGTTCGATCTGCTGCCCCGGCTTCCACACAGGGGAACTGGCTCGGTGCGAC	242		
Db	49223	TGCGCGTTCGCGCTTCGCGCTGCCCGACCCCGCGCGCGCGCGCAGCGACGACGACG	49164		
QY	243	GGCAGCCCCAGCGCGGACAGCTCGTTCAGGCTGGCGTCACGACCTC--TCGCGGGTGGG	300		
Db	49163	CGAGCCCCGAGACGGCGCAGGGGCTATCTGTGTGCAAGAGATTCTGCCCGGGGCT	49104		
QY	301	GAGCGGAAACGCGCACGGCTCAGATCCCTGTCACTCGCATCGGCTCAGTCCGGTCTGTC	360		
Db	49103	GGCGGGGCTCCCGCAGGACAGATTCCACATCAGTGTCTCAGTTCGTTGGCGCGGGTTC	49044		
QY	361	CCCTTGGCTTGGAGGATAGCGTTTACGACGACCGGCACACGCGCGCGGGGGCGGGG	420		
Db	49043	TGGGGGCGCGGACAGT--CAGGTACGCGGGCGGGGCCCGGGGCGGGTCCGGCGCAGG	48986		
QY	421	GGGTTTCAGCCGATTCGCTCGATGACAGCGGCTGGGGGTGGGGCGGTTCGGGCGGATC	480		
Db	48985	GGGCGTGGGCGGGGGGTTCGGGGGCGGGGAGCCGGGGCGGGCGGAGCGCGGGGAC	48926		
QY	481	CGTACCGCCCGGACCGGCTTCGGCC	504		
Db	48925	CGGGACCGGACCGGCGAAAGCC	48902		

RESULT 14  
NSP561198  
LOCUS  
DEFINITION  
NGP561198 89153 bp DNA linear BCT 23-OCT-2003  
Actinomadura sp. ATCC 39727 gene cluster for biosynthesis of  
glycopeptide antibiotic A40926, strain ATCC 39727.  
AJ561198  
ACCESSION  
VERSION  
KEYWORDS  
GI:32487222  
ABC transporter; acyltransferase; beta-hydroxylase; dalbavancin;  
dbv1; dbv10; dbv11; dbv12; dbv13; dbv14; dbv15; dbv16; dbv17;  
dbv18; dbv19; dbv2; dbv20; dbv21; dbv22; dbv23; dbv24; dbv25;  
dbv26; dbv27; dbv28; dbv29; dbv3; dbv30; dbv31; dbv32; dbv33;  
dbv34; dbv35; dbv36; dbv37; dbv4; dbv5; dbv6; dbv7; dbv8; dbv9;  
DcpA protein; DcpB protein; DcpG protein; DcpH protein; DcpI protein;

glycosyltransferase; halogenase; hexose oxidase; HpgT  
aminotransferase; hydroxymandelate oxidase; hydroxymandelate  
synthase; integral membrane ion antiporter; m enoyltransferase;  
mannosyltransferase; monooxygenase; non-ribosomal peptide  
synthetase; ORF1; ORF10; ORF11; ORF12; ORF13; ORF14; ORF15; ORF16;  
ORF17; ORF18; ORF2; ORF3; ORF4; ORF5; ORF6; ORF7; ORF8; ORF9; OxyA  
monooxygenase; OxyB monooxygenase; oxygenase OxyC; prephenate  
dehydrogenase; regulator LuxR family; regulator StrR family;  
response regulator; sensory kinase; type II thioesterase; VanY-type  
carboxypeptidase

Nonomuraea sp. ATCC 39727  
Nonomuraea sp. ATCC 39727  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptosporangineae; Streptosporangiaceae; Nonomuraea.

1  
Sosis, M., Stinchi, S., Beltrametti, F., Lazzarini, A. and Donadio, S.  
The gene cluster for the biosynthesis of the glycopeptide  
antibiotic A40926 by nonomuraea species  
Chem. Biol. 10 (6), 541-549 (2003)  
22721464  
12837387

2 (bases 1 to 89153)  
Donadio, S.  
Direct Submission  
Submitted (12-MAY-2003) Donadio S., Microbial Technologies, Vicuron  
Pharmaceuticals, via R. Lepetit 34, 21040 Gerenzano, ITALY

Location/Qualifiers  
1. .89153  
/organism="Nonomuraea sp. ATCC 39727"  
/mol\_type="genomic DNA"  
/strain="ATCC 39727"  
/db\_xref="taxon:93944"  
/notes="(Actinomadura) sp. ATCC 39727 = cf. Nonomuraea sp."  
complement(x1..141)  
/function="putative Ser/Thr protein kinase"  
/note="ORF1  
outside dalbavancin cluster"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAD91190.1"  
/db\_xref="GI:32487223"  
/translation="MGGSALLPGDPLRGLGSGGGVVVAYDEGRRVAV

CDs

SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

FEATURES  
source

CDS

complement (347.. 1207)  
/function="putative hydrolase"  
/note="ORF2  
outside galbavancin cluster"  
codon\_start=1  
transl\_table=11  
product="hypothetical protein"  
protein\_id="CAD91191.1"  
db\_xref="GI:32487324"  
translations="MRHTVIOQLPVYREAGDPANPLVLLGPPSSAMFRNLL  
ADUASLHLIAPHVFGQSAAPSEVEFPFKLFTILLELDLTINDRPAFYIHDY  
GAPIGLRIASRHPERSAIIQSGNAYLDGFTFWELFAHKDRNTINRPAVELLTA  
DAYRWYTHGVPAQSWLDLSDQAYLDRPGNKVEQLQFWDYQFNLDGYPAPQ  
EYFRFTGPPVYLVAWGANDETFGAAGAEAYARDLPDAEIHLLDAGHFALETHGREIAGI  
IRDFLRRRLRP"  
complement (1797.. 2627)

CDS

```

/functions="putative DNA binding protein"
/note="ORF3
outside dalbavancin cluster"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD91192.1"
/db_xref="GI:32487325"
/translation="MSRQDELREFLRRSRVAPRDVGLPARAGRRVRLREELAA
AAGVSADYYTRMGQGNVSDVLAAGVLRLEAEQRLRELIGDRHHRPPRA
RPAARLEMLDMLGPAALMTGVLDIALNRMGALLTDFEAMPREHRIEWFID
PRVGVDPDWEVAVGAAMALRGALRGSGDDPRPQVLVDSLSARSADFAAHWADHVT
CTYGTGKRIPEHVSQVGMVMTNWESFAPDAFDEVLVYVTAPTGSPSEERLAALAGGVVS

```



```

CDS
    complement(3225..4181)
    /function="putative regulator"
    /note="ORF4"
    outside dalbavancin cluster
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAD91193.1"
    /db_xref="GI:32487226"
    /translation="WNIEDLRETLSEMADEBRRPAPSRFLQGRSLQAREHRRPRRGL
    ALMAAGVVALAVGATFVRGLSPGTSGEAGQPDPAVASVVEQRPALTIVPPGMRL
    AQLLRSLTPVPEFERAARDGAALGLPPYAKRLEGFAAPGSGYEISPAWPGGI
    LQMTTGFRLAEETGLVAGACRGTPLEITLTIASIVQAESFRPEWPKIARLHNR
    LDKWRIQVDSVLYGLKGYSAATPEVRSPPNTVRRPGLPPGPIGSPGADAVRA
    ALEPAGRWLYVADPTGAKFATTQTEVAALVEERDQ"
    complement(4176..4690)
    /note="ORF5"
    outside dalbavancin cluster
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAD91194.1"
    /db_xref="GI:32487227"
    /translation="MDPGEAFAFRDFVGRSGALFRYALITGERAAEDLVQSALAK
    TAAKRGRLDPAAIEGYVRVMYHEQVSWRRRSRAVSSTGWLDPDQARDGHADGL
    RYVMAAALGRLPFRQRTILVRYFEDLSETEIARVILGRVGSVRSQIHRSLRKQTA
    PELTINRFG"
    complement(4620)
    /note="ORF6"
    outside dalbavancin cluster
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAD91195.1"
    /db_xref="GI:32487228"
    /translation="WAMPRIIFGLENYXICTFKGORRLSPDEVARYLFRKWSWGR
    SNVFLNGCARLYLVGSHPEYATPCDNVVELVTHDKAGRIEGLVDAERLREE
    GIAGDYLKNNIDSNAGSYGCHENTCVRHGFGRGLADVLFLVTVQVCGAGKVL
    QPFRVAYCLSRAGHIEWGVSSATRSRPIINTEDEPHGDAERPRRLHVIUGSNMS
    ETTLLKIGATDLVLMIEAGVYMRDLSENFIRAIRREVSHDMTGRKVRILNGEAS
    ALEIOEYLSKAADFVRRGDATLRVLEWERTLRAVETGDLVLRGEIDWTKYQ
    MIERVSKYDLPDLSRSVAQLDAYHDVHRRGLFYLQRLKGLVERVASDVKIFEAKS
    VPPQTTARLRGEFIRKAQEKRRDFTVDVHVLKLNDAQRTVLCKDPRFSVDVRL
    IAGM"
    complement(6242..7342)
    /gene="dbv1"
    complement(6242..7342)
    /gene="dbv1"
    /codon_start=1
    /transl_table=11
    /product="putative hydroxymandelate oxidase"
    /protein_id="CAD91196.1"
    /db_xref="GI:32487229"
    /translation="WHESPVCLAEYEEITAAKYLPAVDRDFIDGSGRSGQTLRANRAAF
    DRVFLVRLQVDSACTRLLGHFATPVAVAPVAVVHRLVHPDGEIATARAADAG
    PFTVTSLSVPEVDVETALGHWVWFOLYCLREHAATLGLIRAEADAGCRALMLTLDVP
    WNGRRPRDIRNRPFLPPHYRPHVLTANSCTEAHAGSAGSALAAHTAMLSAAMDVSY
    LETLRASGLPVVVGILHPEDABRADLIDIGIVVSNEGGRQLDGAVASLDALPGVA
    ESVGRCETMLDGGVRSQADVLKALALGASGLVIRPVINGWLAADGERGVRVILGLG
    AEIEDGLGAGCDVAAQAALRTFRPAGFVS"
    complement(7461..8531)
    /gene="dbv2"
    complement(7461..8531)
    /gene="dbv2"
    /codon_start=1
    /transl_table=11
    /product="putative hydroxymandelate synthase"
    /protein_id="CAD91197.1"
    /db_xref="GI:32487230"
    /translation="MESLPPPLAVYVMYVADLKVATLPWTDEYRFVAVVGTASDHR
    SVALRGRITLVLTQSDGHPASAVVRTHGDVQIALRTPDVDVVFTHAAGARP
    VRSFSPHGGPGPACSAIIGFGVDVHTLVQRDPDPPGLPVGFSEAPSAESAEGDAAE
```

```

gene
    complement(8760..11363)
    /gene="dbv3"
    /gene="dbv3"
    /codon_start=1
    /transl_table=11
    /product="putative regulator LuxR family"
    /protein_id="CAD91198.1"
    /db_xref="GI:32487231"
    /translation="MLFGRDELKSLTRLLDSTAAGRGVAVITGPVVGKTAHLHEL
    GMSIAAGRLTARTCPAEPQSLDNGVADQILGRGAERLTARRGDAVEDVCVSLFQ
    MEANFILLITDDVDLADDFSLAILSMTPLLTDRMTAVITICQRPAPPLPHVAES
    LELPGIIEVELPLPPAVROFATEHGAETAADLADLLYRSGSPILLVRLIEDQ
    EAGAGLVVDGSPFMSAVAAVCGCEPEAVVAEAVLGEHATPDVAGELVGTAPPA
    TRSMGMLERAGLLAGGRFRHEAGRLAVLGWTSYGEWELIRRAAEILHRRGGPPSAVA
    TRLEAGWSGEWEAFDVLVEAGQAQDEGDFVAMKCLRLALASGWTGTPRLLDVKVWL
    AAAEWRVDPAAARHVPDLDLATSGLARGSGHMLFRLQWYGRFADAAELIDRLRP
    SVADRDAASLIANCHVPLVLDRLPRSGMSQTVEDARRILRQAETDEAMD51I
    SALMALLGGVSEVAASCETLLKPGVTKAPTWKAIISAIRAETAWRKGLAGAEHA
    OEALITLQPSGWGVAIGAPLSTLLHAQTAMGLDEAKATVAVMPRETATATGIGYE
    LARAHVLTQEPRAAFACGLACQAVQRMGSSLSDVVPWRLCAAEACIOLGWRRAA
    DLVTAIHTSSGLTITYGVALRLHAQLSKPAQRORLLQSVSDALAAQDRIQLALS
    CDLAGTFQLKGGDEARAYVRAQELARECNAPLMWRRLAAQHDHGETAFPSGAERV
    AVLAARGTNRREAEALYITRSTVEQHLTRIYKHLHVQTRGDLGNLFAADIADKATAT
    AGREPRAVRL"
    complement(11468..12433)
    /gene="dbv4"
    complement(11468..12433)
    /gene="dbv4"
    /codon_start=1
    /transl_table=11
    /product="putative regulator StrR family"
    /protein_id="CAD91199.1"
    /db_xref="GI:32487232"
    /translation="MDPTGVDIATLPWRIELSRLSVVSYPRTSGDPHVEHTLSAQ
    GEUPLPILVHPTWRVIDGLHRLKVARVRGETTISRLLIDGTSDDAFVLAENVRGIL
    PLSADKRAAVRIIGTHPQWDRKRVASATGISAGVATDURRRGGOGGDEARLGRDGR
    IRPVDSEGRRLAAELIRSHPDLSLRQVAKQVGISPETVNDVRGRLEHGESP1PDGSR
    RLRTKPELRAEQDFGHVDRDQAVLERLKADPALRLTETGRILLMLSLHSIDQG
```

```

Query Match      10.7%; Score 55.6; DB 1; Length 89153;
Best Local Similarity 48.7%; Pred. No. 15;
Matches 151; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

Qy      6  AGCAGAAACGACCGGTGGTCCGCCGCCCATACGAACTGCCACTCCGGCGGGGTG 65
Db      32170 AGCAGCCCAACCCCGCTCTCACGGCGCACGCCGCCACCCAGCAGCCGCGCAGCTCC 32229

Qy      66  ACCAGGTGACAGGTGGAGTTGAGCAGCAGCTCATCCCGCTGCGCGCGGCATCGC 125
Db      32230 TCCGCGCTGTCACAGATCCGCTGTACAGAGGCTTCGGGCGCGGTCCACACCGCACC 32289

Qy      126  AAACACCGCGGCGAGGATCAGTGCAGCAGCGCGATCGCGCTCGATCTCGACGGTCCGG 185
Db      32290 GCATCCGCGGAGCGCTCCACTCGCGCGGAAACAGCATCGCACCGGCTCCGCGCGGGC 32349

Qy      186  CGCAGCTCGATCTCGTCCGCCGCTCCACACGAGGGAACTGGCTCGGTGGGCGAGCGC 245
Db      32350 GGCACGCGGCTCTGTTCCTCCACTCTCCACACGAGGGCGGCGCTGCTCAGGACCGATCAGG 32409

Qy      246  AGCCCCACCGCGGACAGCTCGTCCAGGCTGGCGTCAACGAGACCTCTCCGCGGTCCGGAGCG 305
Db      32410 CCCACGCGCGCGACCGCGCGCGGCTCGGCCACCCACTGCTCCAGTCCGCGGAGGATC 32469

Qy      306  GAAACGGCGCA 315
Db      32470 GACGCGAGCA 32479
```

```

RESULT 15
SC0939131 303550 bp DNA linear BCT 11-FEB-2003
LOCUS Streptomyces coelicolor A3(2) complete genome; segment 28/29.
DEFINITION AL3939131 AL158067 AL158060 AL158061 AL356592
ACCESSION AL356932 AL357613 AL590435 AL592126 AL596248 AL645882
VERSION AL3939131.1 GI:24418961
KEYWORDS
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1 Bentley, S.D., Chater, K.F., Cerdano-Tarraga, A.M., Challis, G.L.,
Thomson, N.R., James, K.D., Harris, D.B., Quail, M.A., Kieser, H.,
Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,
Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,
Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,
Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S.,
Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,
Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrall, B.G.,
Parkhill, J. and Hopwood, D.A.
Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
NATURE 417 (6885), 141-147 (2002)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21996410
PUBMED 12000953
REFERENCE 2 (bases 1 to 303550)
Bentley, S.D.
Direct Submission
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: sd@sanger.ac.uk
ON or before Oct 29, 2002 this sequence version replaced
GI:20520692, GI:20520799, GI:20520800, GI:20520856, GI:20520802,
GI:20520783, GI:20520865, GI:20520870, GI:20520878.
LOCATION/Qualifiers
1. 303550
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
135..1643
/gene="SC07252"
/note="synonym: SC5H1.40c"
135..1643
/gene="SC07252"
/note="SC5H1.40c, possible regulatory protein, len: 502
aa; similar to e.g. SW:SP15_STRGR (EMBL:M32687)
Streptomyces griseus sporulation protein (529 aa), fasta
scores; opt: 394 z-score: 452.2 E(): 7.2e-18, 27.9%
identity in 530 aa. Similar to SC7A1.26 (EMBL:AL034447)
S.coelicolor possible transcriptional regulator (500 aa)
(43.0% identity in 491 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="CAB42965.1"
/db_xref="GI:4835343"
/db_xref="SPITREMBL:Q9X7U7"
/translation="MSKPNRLSLDFGLAGNSKGELARLVNPNQAAAGHPQLADTDS
RVRWIDMIPRPDPVPRVLAALFTERGLVVTIEDLGLVHGRAGKRPDSEEPD
GPWPAPRTAAVTFTFGMDLNNRGLVGAGLAAGSALSAMDHLHTPALAAD
APDLQPLHADPAGDFDYEAPITGSOEVELERSVEFWAADAAGGLQKRAVGQL
NEVGMALVHPHCHLORLAVANLAVLAWNSHDVGLPELTACKYFVIAAAREGG
DRPAGCALSRAROMYHLGKPDALDLMLKLAGSGEQLVPTKMLTITENWAQAS
MGKQAVRTIGRAEDLUFVSKADVPFPDWMQTFKEDDLYGMOALAIRLAEFPGAA
AQHAYKALALIDGRQSKIFQLNSMASACFIADDPQADRYARLALVNSGNS
QRTWDLRQMYRTAEYAGPKYQELREELIKALPKGSPRGPGGTAPA"
complement (1663) .2979)
/gene="SC07253"
/note="synonym: SC5H1.39"
complement (1663) .2979)
gene
CDS
/gene="SC07253"
/note="SC5H1.39, hypothetical protein, len: 364 aa;
unknown function, similar to TR:P73120 (EMBL:D90903)
Synchocystis sp, hypothetical protein (358 aa), fasta
scores; opt: 492 z-score: 559.6 E(): 7.6e-24, 31.6%
identity in 364 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB42962.1"
/db_xref="GI:4835340"
/db_xref="SPITREMBL:Q9X7U4"

```

```

/translation="MEHDLSPATLGLRPRPVAVSVLTTHRRPDPNQRVRLRN
AVAAKQLEDDPAVSRDRAAEVSLDLRALAEIDLTYAEDGLAIFAAPGEHQWTL
RSVPVSTDTFRNIAVAQAARPFVWLSVADRVKLSGAGRVVHEHLGGFPL
DRPDDFQERQSTFADPSTFDEGTFRVFLRDAAAGRIILRAHPRFLVYTGPEPL
SLFERAGALADAVLVPHGLAHGTAARVWQAVPVLKEARGSVAVARELDAARGR
KDFAGVDELAEASHTGRVLLAVENPKVTRMDGHELLPADDDLDAREDIVDEIV
EQCLTGADVRFVDPGALGDVVGAGVRLY"
complement(5513..5516)
RBS
/notes="possible RBS"
gene
complement(5580..7211)
/genes="SCO7256"
/notes="synonym: SC5H1.36"
CDS
complement(5580..7211)
/genes="SCO7256"
/notes="SC5H1.36, probable protease (putative secreted
protein), len: 543 aa; similar to many e.g. TR:Q54398
[EMBL:L42758], SlpD, Streptomyces lividans protease, (539
aa), fasta scores; opt: 697 z-score: 769.7 E(): 0.35.8%
identity in 545 aa overlap. Similar to other S.coelicolor
proteases e.g. SlpD, SCH5.02c [EMBL:AL03563] proteinase
(539 aa) (35.8% identity in 545 aa overlap). Contains
possible N-terminal signal sequence. Contains pfam match
to entry PF00561 abhydrolase, alpha/beta hydrolase fold,
score 55.00, E-value 1.6e-12"
/codon_start=1
/transl_table=11
/product="putative protease (putative secreted protein)"
/protein_id="CAB42961.1"
/db_xref="GI:4835339"
/db_xref="GOA:Q9X7U3"
/db_xref="SPTREMBL:Q9X7U3"
/translation="MQAKLETRPSVRCVAVTGALGLALLGALPAAGATTATTRIT
RGSLSRFYQVETWACETEMPKDLOQKVTVPLDYSRPGRGTLVALARYATGTS
GSLVLLNFGSGGSGINELALGKGFHELTNGVDVTFDPDGVGRSSPVSCGPATLKI
MEATDGGSTDPGVLERLDAARCAKVSGLVLEHITGVDAARDVDVWRALGDDR
LVNLYGESVGLRGLAVYARFPDKVGRWLVLDGYDLTRPLEACGLAGRGQTALENFL
DWCVEDVACFGQDARDQVERLVASLSDSPVSAPFEFTQDQWVGAIQALYER
LNLPEALAIACQLEDGDPRTAAQVTSGLRLARYEDVSPVFGRIYRLTVLMCYGRPR
GTYIDRDVDLDVLRMLLVGRGDPATPYRTWTETADRLGPSAVLNDNRGEHTGYA
SSKCVHRKVDDFLVGLSLPPDGSGCPPESTGDGSG"
complement(5910..6809)
misc_feature
/genes="SCO7256"
/notes="Pfam match to entry PF00561 abhydrolase, alpha/beta
hydrolase fold, score 55.00, E-value 1.6e-12"
complement(7217..7220)
RBS
/notes="possible RBS"
7437..7440
RBS
/notes="possible RBS"
7445..8158
gene
/genes="SCO7257"
/notes="synonym: SC5H1.35c"
CDS
7445..8158
/genes="SCO7257"
/notes="SC5H1.35c, possible secreted protein, len: 237 aa;
unknown function, probable CDS suggested by positional
Query Match 10.7%; Score 55.6; DB 1; Length 303550;
Best Local Similarity 48.4%; Pred. No. 11;
Matches 246; Conservative 0; Mismatches 254; Indels 8; Gaps 3;
QY 8 CACGAAACACCGGTGTGTCGCCCGCGCCGATACAGAACTGCCACTCCGCGGGGTGAC 67
DB 72105 CACGTCCGGGCGCATAGCGCCAGCGGGCGGCTTCGCGCGGGAACCGGGCTTCC 72164
QY 68 CAGGTGACAGGTGGAGTGGAGCAGCGCAGCTATCGCGCGCTGCGCGGCGGCGGAA 127
DB 72165 CACCAAGGTGAGGAGCCCGTGTACCGCGCGGACAGATGGCCACCGCGCGCGGAG 72224
QY 128 ACACCGGGCGAGGATCACGTGCGAGCAGCGCGATGCGCGCTTCGTCTCGAGGTGCGCG 187
DB 72225 CCGGGGCGAGCGCTCTCAGCAGCGCGGTGCGCGAGCGGCTTGCGTGTCTCGCGGCGCAG 72284

```

```

QY 188 CAGCTCGATCTCGTCCGCCCGGTCCACACAGGGGAAACTGCTCGTGGCAGCGCGCAG 247
DB 72285 CGCGCGCTCTCGCGCGCGCGCGAGGCGTACGGCGCGCGACTCGGCCACCGCTGTGCGCC 72344
QY 248 CCCACAGCCGGGACAGCTCGTCCAGGCTGCGTCAAGGACCTCTCGCGGGTTCGGAGCGGA 307
DB 72345 CTGCGTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 72404
QY 308 AACCGCGACGGGTTCAGATCCCTGTTCAGTCGTCATCGGCTCAGTCCCGGTTCGTCCTCCCTTGG 367
DB 72405 GGCGCGACAGCGCGCGGTTCGCGCGCA--GAAACGGCTCCGTCACCTCCGACATGCGGTC 72461
QY 368 CTGGGAGGATAGCGGTTTACGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 427
DB 72462 GCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 72521
QY 428 AGCC-GATCGCTCGATGATGACCGCGGCTCGGGGTTCGGGCGCG--TGGGCGAGATCCG 482
DB 72522 GGCAGCTCGCGAGCTGCGCGGCGGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72581
QY 483 TACCGCGCGGACCGCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 510
DB 72582 GACGCGCTTGGCGAGCTCGATCAGCGGCC 72609

RESULT 16
AB067802S1
LOCUS Homo sapiens gene for RNA binding protein HOK, exon 1. 780 bp DNA linear PRI 10-OCT-2001
ACCESSION AB067802
VERSION AB067802.1 GI:15991315
KEYWORDS 1 of 8
SEGMENT Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Li,Z.Z., Kondo,T., Murata,T., Ebersole,T.A., Nishi,T., Ushio,Y.,
Yamamura,K. and Abe,K.
TITLE Expression of a KH RNA binding protein, Hqk, is altered in human
glioma
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 780)
AUTHORS Li,Z.Z. and Abe,K.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-2001) Kuniya Abe, Institute of Molecular
Embryology and Genetics, Kumamoto University; Kuhnouji 4-24-1,
Kumamoto, Kumamoto 862, Japan (E-mail:kabe@go.kumamoto-u.ac.jp,
Tel:81-96-373-6597(ex.6597), Fax:81-96-373-6597)
FEATURES
source
1..780
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q25-q27"
/tissue_type="brain"
<552..693
/genes="Hqk"
/product="RNA binding protein HOK"
/notes="CDS is reported in Acc#:AB067808"
/number=1
exon
ORIGIN
Query Match 10.6%; Score 55.2; DB 9; Length 780;
Best Local Similarity 48.4%; Pred. No. 55;
Matches 183; Conservative 0; Mismatches 193; Indels 2; Gaps 1;
QY 6 AGCAGCAGACCGGTGTGTCGCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 65
DB 141 AGCCCCGGCACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 200

```

QY 66 ACCAGGTCGACCGAGTGGAGTTGAGCAGCAGCTCATCGCCGCTGCGCCGCGATGCGG 125  
Db 201 CCGGGTCTCTCTCTCGTCCGCGCGCAGCAGCGCCGCGCGCGCCGCACTCGCGCCAGCG 260  
QY 126 ABAACCGGCGCGAGTACGTCGACGAGCGGATGCGGCGCTCTCGATCTCGACGCTCGGC 185  
Db 261 AGACCCGCTCCGCGCGCGCGCGCTCTGAGGAGCGCGCGCGCGCGCGCGCGCGAG 320  
QY 186 CGCAGCTCGATCTGTCGCGCGCGCTCCACACAGCGGGAATGCTCGGTGCGCAGCGC 245  
Db 321 AGCGCGCGCGCTGGGAGCGC--TCGGGCGCGCGGAAAGTCTGCTGCGCGCGCGCGCG 378  
QY 246 AGCCCGAGCGGACGCTGTCAGGCTGCGCTGCGGACCTCTCGCGGTCGCGGAGCG 305  
Db 379 AAGCGCGGCTGCGCGCGCGCGCGCTGCGCGCGGAGCGGAGCGCGCGCGCGCGCG 438  
QY 306 GAAACGCGCGCGCTCAGATCCCTGTGTCGATCGGCTCACTGCGGCTGCTGCGCCCTT 365  
Db 439 GAGCGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 498  
QY 366 GGCTGGGAGATAGCGG 383  
Db 499 CTCGCGCGCGCGCGCGG 516  
RESULT 17  
ID BX842572/c standard; circular genomic DNA; PRO; 341957 BP.  
XX AC BX842572; AL021428; AL021926; AL021927; AL021928; AL021929;  
AC AL021930; AL123456; Z74410; Z80233; Z80775; Z86089; Z92669; Z92770; Z96071;  
AC Z97050;  
SV BX842572.1  
XX 21-NOV-2003 (Rel. 77, Created)  
DT 21-NOV-2003 (Rel. 77, Last updated, Version 1)  
XX  
DE Mycobacterium tuberculosis H37Rv complete genome; segment 1/13  
KW complete genome.  
XX  
OS Mycobacterium tuberculosis H37Rv  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.  
XX  
EN [1]  
EX MEDLINE; 98295987.  
RX PUBMED; 9634230.  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eiglmeier K., Gas S., Barry III C.E., Tekala F., Badcock K.,  
Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,  
Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K.,  
Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J.,  
Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,  
Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,  
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete  
genome sequence";  
RL Nature 393:537-544(1998).  
XX  
XX [2]  
RX PUBMED; 12368430.  
RA Camus J.C., Pryor M.J., Medigue C., Cole S.T.;  
RT "Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv";  
RL Microbiology 148:2967-2973 (2002).  
XX  
XX [3]  
RP 1-341957  
RT Parkhill J.;  
RL Submitted (11-JUN-1998) to the EMBL/GenBank/DBSJ databases.  
RL Submitted on behalf of the Mycobacterium tuberculosis sequencing and

mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,  
Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut  
Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail:  
parkhill@sanger.ac.uk  
XX Notes:  
CC Details of M. tuberculosis sequencing at the Sanger Centre  
CC are available on the World Wide Web.  
CC (URL, http://www.sanger.ac.uk/Projects/M\_tuberculosis/)  
XX  
FH Key Location/Qualifiers  
FT 1. 341957  
FT /db\_xref=taxon:83332"  
FT /mol\_type="genomic DNA"  
FT /organism="Mycobacterium tuberculosis H37Rv"  
FT /strain="H37Rv"  
FT 1. 1524  
FT /evidence=EXPERIMENTAL  
FT /note="RV0001. (MT0001, MT0029.01, P49993), len: 507 aa.  
FT dnaA, chromosomal replication initiator protein (see  
FT citations below), equivalent to other Mycobacterial  
FT CHROMOSOMAL REPLICATION INITIATOR PROTEINS e.g.  
FT P46388|DnaA MYC from Mycobacterium leprae (502 aa);  
FT Q9L7L7|DnaA MYC from Mycobacterium paratuberculosis (509  
FT aa); P49990|DnaA MYC from Mycobacterium avium (508 aa);  
FT P49992|DnaA MYC from Mycobacterium smegmatis (504 aa);  
FT etc. Also highly similar to others except in N-terminus  
FT e.g. Q9ZHT5|DnaA\_STRCH CHROMOSOMAL REPLICATION INITIATOR  
FT PROTEIN from Streptomyces chrysomallus (624 aa);  
FT Q9ZHT6|DnaA\_STRRE from Streptomyces reticuli (643 aa);  
FT DnaA\_ECOLI|P03004|B3702 chromosomal replication initiator  
FT protein from Escherichia coli strain K12 (467 aa), FASTA  
FT scores: opt: 986, E(): 0, (43.2% identity in 389 aa  
FT overlap); etc. Contains PS00017 ATP/GTP-binding site motif  
FT A (P-loop) and PS01008 dnaA protein signature. BELONGS TO  
FT THE DnaA FAMILY. Note that the first base of this gene has  
FT been taken as base 1 of the Mycobacterium tuberculosis  
FT H37Rv genomic sequence."  
FT /transl\_table=1  
FT /function="PLAYS AN IMPORTANT ROLE IN THE INITIATION AND  
FT REGULATION OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN  
FT OF REPLICATION. IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA  
FT AT A 9 BP CONSENSUS (DnaA BOX): 5'-TTATC(C/A)C(A/A)-3'.  
FT DnaA BINDS TO ATP AND TO ACIDIC PHOSPHOLIPIDS. DnaA PROTEIN  
FT BINDS THE ORIGIN OF REPLICATION (oriC), ATP AND ADP, AND  
FT EXHIBITS WEAK ATPase ACTIVITY."  
FT /gene="dnaA"  
FT /locus\_tag="RV00001"  
FT /product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DnaA"  
FT /protein\_id="CAA16238.1"  
FT /translation="MIDDPGSGFTVWVAVSELNGPKVDGSSDNLASPLTPQOR  
FT ARLNLVQLTIVEGFAALLSVSPFVNEIHLRAPITDLSRLGHQIQGVRIAPPA  
FT TDEADTTVPSENPATSPDITDNDIIDAARAGNQHSWPSYFTFRPHNTDSATA  
FT GVTSLNRYTFTFVIGASRFAPAALAIAPAPAYNPLFIWGESGLGTHLHAAG  
FT NYAQLFFPMVKVYVSTEEFTNDINSLDRKVKRSYRVDVLLVDDVDFTEGKEG  
FT IQEPPHFTLNANKOIVSSDRPPKQATLDRLRTFEGWLTIDVQPELETRIA  
FT ILRKAQERLAVPDVLELIASITERNRELEGALJVTAFASINKTPIDKALAEIVL  
FT RLIDADANTMQISAATMAATAEFTTVEELRPGKTRALLAQSRQIAMIYLCRELTDL  
FT LFKIGQAFGRDHTTVMYAKRILSEMAERREVFHVXELTTRIKRQR"  
FT  
FT misc\_feature  
FT /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
FT 1384..1440  
FT /note="PS01008 DnaA protein signature"  
FT 2052..3260  
FT  
FT misc\_feature  
FT /evidence=EXPERIMENTAL  
FT /note="RV0002, (MTV029.02, MTCY10H4.0), len: 402 aa. dnaN,  
FT DNA polymerase III (beta chain) (EC 2.7.7.7) (see citations  
FT below), equivalent to other Mycobacterial DNA POLYMERASES  
FT III BETA CHAIN e.g. NP\_301130.1|NC\_002677 from  
FT Mycobacterium leprae (399 aa); Q9L7L6|DP3B\_MYC from  
FT Mycobacterium avium subsp. paratuberculosis (399 aa);  
FT P52851|DP3B\_MYC from Mycobacterium smegmatis (397 aa);





MEDLINE 22595073  
FUBMED 12640520  
REFERENCE 2 (bases 1 to 716)  
AUTHORS Danchin, A. and Pascal, G.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,  
Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong  
FEATURES Location/Qualifiers  
1..716  
/organism="Penicillium marneffei"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:37727"  
/clone="pml2h3.g"

ORIGIN  
Query Match 10.5%; Score 54.8; DB 11; Length 716;  
Best Local Similarity 47.0%; Pred. No. 64;  
Matches 207; Conservative 0; Mismatches 227; Indels 6; Gaps 1;

Qy 77 CAGTGGAGTTGAGCAGCCAGCTATCGCGCGCTGCGCGCGCATGCGGAAACACCGCGC 136  
Db 638 CGGCGCGCAGCG 579  
Qy 137 CAGGATCAGTGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 196  
Db 578 GAGCGGGGGGCAAGCG 519  
Qy 197 CTGCTGCGCGCGCGTCCACACAGCGGAACTGCTCGTGGGAGCGCGCGCGCGCGCGCG 256  
Db 518 CTCGCCCG 459  
Qy 257 GCACAGCTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 316  
Db 458 CGGCG 399  
Qy 317 GCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 376  
Db 398 GCG-----CG 345  
Qy 377 ATAGCGTTCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 436  
Db 344 CG 285  
Qy 437 GCTGATGACAGCGCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496  
Db 284 CG 225  
Qy 497 CTTGCG 516  
Db 224 CCACCG 205

RESULT 21  
AF288517 4283 bp DNA linear PRI 02-AUG-2001  
LOCUS Homo sapiens mutant insulin receptor substrate 2 (IRS2) gene,  
partial cds.  
ACCESSION AF288517  
VERSION AF288517.1 GI:15077067  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 4283)  
AUTHORS Heyne, B., Gehrisch, S. and Jaross, W.  
TITLE Two missense mutations in insulin receptor substrate 2 (G879S and G882A)  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4283)  
AUTHORS Heyne, B., Gehrisch, S. and Jaross, W.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUL-2000) Institut fuer Klinische Chemie und

Laboratoriumsmedizin, Universitaetsklinikum der Technischen  
Universitaet Dresden, Fetscherstr. 74, Dresden 01307, Germany  
FEATURES Location/Qualifiers  
1..4283  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="13q34"  
/note="obtained from a patient with diabetes mellitus"  
139..>4283  
/gene="IRS2"  
4139..>4150  
/gene="IRS2"  
/product="mutant insulin receptor substrate 2"  
139..>4150  
/gene="IRS2"  
/note="similar to NM\_003749"  
/codon\_start=1  
/product="mutant insulin receptor substrate 2"  
/protein\_id="AAK81053.1"  
/db\_xref="GI:15077068"  
/translation="MSPRRHGPFGPAGSDGPNLNNNNNNNNHNSVRKQYLRKQKHGH  
KRFFVLRGAGGDEATAGGSAPOPRLVEYSEKWRKAGAPKRVIALDCCCLININ  
KXADAKHYLIATYTKDRYFAVAENEQEQWYRALTDLVSEGRAGADAPAAAPA  
ASCSASLPGALGGSAGAGADSGLVAPATAAEVYQVNLKPKGLGSKNLTGVYR  
LCLSAITGFKVLCNQPSVTQLQNMIRCHSDSPFTEVGRSAVTGFGELWQADD  
SVVAGNIHETILEAMKULKEFPRKSKSSSSSSATGPIISVPGARHHLNLDPPS  
QTGLVRRERTDLAATPPAAKSCSRVTSSEGGGAAAGAAAGAAAGAAAGAAAGAA  
GVPAPLRSHTLSGGCGRGSKVALLPAGALQHSRSMFVAHSPPAATSPGLSLP  
SSGCGSGVPPPPPHPLPHLPQHPQRPFGSGSASASGSPSDGFMGLDEYSGSPG  
DLRACSHRSNTPEIAETPPADGGGGGFEVGYMTDRPLSHGCRSVRRVSGDAADQ  
LDGRKRTYSLTTPAROPVPOPSASLDEYTLMTATFSGSAGRLCPSPASSPKVA  
YHPYEDYDIEIGHSRSSNLGADDMPTFGAALAGSGSGSCRDDYMPKSPAS  
VAPKQILQPRAAAAAASFAGPAGPAPTSAGRTFPAGCGKASAPAESFEDS  
GSRMWCCKSKSMEHADKLLPNVDYLVNSPDAVTTPPPDFFAALHPGGEPLRGV  
PCCYSILPSYKAPYTCGSDSDYQLMSSPVRIEERLEPQATPGSQASAFGA  
GCTQPHVPVSPVPSGSDPSFLAQGRVPRTRLESLPLSPMHEVPLPPEPK  
SPGVINDFGSPGARLGPAPPLASASSSLLSASSPSSLGSTGTPGSSDRQK  
SFLSYMLDDESSPKSPKPGHPVGLDGLLSPEASSVPPLPPRPSAPSSSLQ  
PPPPAPGELYRLPPASAVATAQFGAASSLSDTGNDGYTEWAFVATPPPIA  
APFKPAKAVTSVSGVRLSLMEQVSGVEAFLOASQPPDPHRAKYIRADPQGRRR  
HSSETFSSITTVFVPSFAHNPKHNSASVENLSKSSGSGVGGVGGGGDEPTSP  
RGLQAPAPLAPQREWTQFGGLGCGSGSPMRRTSAGFQNGLYAIADVREP  
GLPPQPPPPPLPQPGDKSMGRSLGLISAVGVSGTGGCGCGGPGPGLPPANTY  
ASIDFLSHLKEATVK"  
2773  
/gene="IRS2"  
/note="compared to IRS2 deposited in GenBank Accession  
Number AF322114; mutant encodes serine instead of wildtype  
glycine"  
/replace="g"  
2783  
/gene="IRS2"  
/note="compared to IRS2 deposited in GenBank Accession  
Number AF322114; mutant encodes alanine instead of  
wildtype glycine"  
/replace="g"

variation  
variation

ORIGIN  
Query Match 10.5%; Score 54.8; DB 9; Length 4283;  
Best Local Similarity 46.2%; Pred. No. 41;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
Qy 46 ACTGCCACTCCCGCGGGGTGACCCAGGTGCGAGGTGGAGTTGACACCGAGCTCATCG 105  
Db 1034 AGTTCCGGCGCGCAGTAAAGACCAATCTCGGGGTCTGTCGGCCACGCCATCAGCG 1093  
Qy 106 CGCGCTGCCCGCATGCCGAACCGCGCGCGAGATCATCGTCAGCAGCGGATCGCG 165  
Db 1094 TCCCG 1153  
Qy 166 CCGCGATCTCGACGCTCGCGCGCGCGCGCGCTCGATCTCTGTCGCCCGCGCTCCCACACCGAGGGAA 225



```

Db 1154 TGGCGCGCTCGCGCACCGACAGCTGGCGCCACCCCGCGCGGCGCAAGTCAGCTCGT 1213
Qy 226 ACTGGCTCGGTGGCGAGCGGAGCCCGACCGCGGAGCAGCTCGTCCAGGCTGGCGTCAAGGA 285
Db 1214 GCGGGTGGCGACCGCGAGCGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1273
Qy 286 CCTCTCGCGGTGGCGAGCGGAAACGCGACCGCTCAGATCCCTGTCAGTCCGATCGGCT 345
Db 1274 GCGCAGCGCGGTGTCGTTGGTGGAGC-CCCTGAGCCCGCGCGCGTGGCGCGCGCC 1332
Qy 346 CAGTGGCGGTGTCCTCCCTGGCTGGAGATAGCGGTTTCAGACGAGCGCGCACACCGG 405
Db 1333 CTGAGCGCTCGCACACCTGAGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCTG 1392
Qy 406 CGGCGCGCGCGCGCGGTTCAGCGGATCGCTCGATGACGACCGCGCTCGCGGCTCGG 465
Db 1393 CTGCGCGCGCGCGCGCTGCAACACGCGCTCCATGTCCTGCGCGCGCGCGCTCG 1452
Qy 466 GCGGTGGCGAGATCCGTACCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCG 515
Db 1453 CGCGCGCGCGCGCGCGCTCCCTGTCGTCAGCAGCGCGCGCGCG 1502

```

```

RESULT 22
AF322114S1
LOCUS Homo sapiens insulin receptor substrate 2 (IRS2) gene, exon 1.
DEFINITION AF322114
ACCESSION AF322114
VERSION AF322114.1 GI:12247738
KEYWORDS
SOURCE 1 of 2
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4283)
AUTHORS Heyne,B.
TITLE Insulin receptor substrate 2 gene sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4283)
AUTHORS Heyne,B., Gehrisch,S. and Jaross,W.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2000) Institut fuer Klinische Chemie und
Laboratoriumsmedizin, Universitaetsklinikum der Technischen
Universitaet Dresden, Fetscherstr. 74, Dresden 01307, Germany
FEATURES
Location/Qualifiers
1..4283
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q34"
/notes="diabetes mellitus patient"
<1..4150
/gene="IRS2"
/number=1

```

```

exon
Query Match 10.5%; Score 54.8; DB 9; Length 4283;
Best Local Similarity 46.2%; Pred. No. 41;
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;

Qy 46 ACTGCCACTCGCGCGGTGACCGAGTCCAGCGGCGAGGATCAGTGCAGCAGCGGATCGCG 105
Db 1034 AGTTCCGCGCGCGAGTAAGACCAATCGTCGCGGTGTCGCGCGCGCACCCCATCAGCG 1093
Qy 106 CCGCTCGCGCGGATCCGAAACACCGGCGAGGATCAGTGCAGCAGCGGATCGCG 165
Db 1094 TCCCGCGCGCGCGCGCACCAACCACTGTGTCACTGCCCCCGCGCGAGACGCGGCTGG 1153
Qy 166 CCTCGATCTCAGCGTGGCGGAGCTCGATCTCTGTCGCGCGCGCTCCACACCGAGGGA 225

```

```

Db 1154 TGGCGCGCTCGCGCACCGACAGCTGGCGCCACCCCGCGCGGCGCAAGTCAGCTCGT 1213
Qy 226 ACTGGCTCGGTGGCGAGCGGAGCCCGACCGCGGAGCAGCTCGTCCAGGCTGGCGTCAAGGA 285
Db 1214 GCGGGTGGCGACCGCGAGCGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1273
Qy 286 CCTCTCGCGGTGGCGAGCGGAAACGCGACCGCTCAGATCCCTGTCAGTCCGATCGGCT 345
Db 1274 GCGCAGCGCGGTGTCGTTGGTGGAGC-CCCTGAGCCCGCGCGCGTGGCGCGCGCC 1332
Qy 346 CAGTGGCGGTGTCCTCCCTGGCTGGAGATAGCGGTTTCAGACGAGCGCGCACACCGG 405
Db 1333 CTGAGCGCTCGCACACCTGAGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCTG 1392
Qy 406 CGGCGCGCGCGCGGTTCAGCGGATCGCTCGATGACGACCGCGCTCGCGGCTCGG 465
Db 1393 CTGCGCGCGCGCGCGCTGCAACACGCGCTCCATGTCCTGCGCGCGCGCGCTCG 1452
Qy 466 GCGGTGGCGAGATCCGTACCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCG 515
Db 1453 CGCGCGCGCGCGCGCGCTCCCTGTCGTCAGCAGCGCGCGCGCG 1502

```

```

RESULT 23
AF385933
LOCUS Homo sapiens insulin receptor substrate 2 insertion mutant (IRS2)
DEFINITION AF385933
ACCESSION AF385933
VERSION AF385933.1 GI:14537856
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4286)
AUTHORS Heyne,B.
TITLE Two insertions in insulin receptor substrate 2 (N28_H29insN;
N28_H29insNN)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4286)
AUTHORS Heyne,B., Gehrisch,S. and Jaross,W.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2001) Institut fuer Klinische Chemie und
Laboratoriumsmedizin, Universitaetsklinikum der Technischen
Universitaet Dresden, Fetscherstr. 74, Dresden 01307, Germany
FEATURES
Location/Qualifiers
1..4286
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q34"
/notes="isolated from a patient with type 2 diabetes mellitus"
<139..>4153
/gene="IRS2"
<139..>4153
/gene="IRS2"
/product="insulin receptor substrate 2 insertion mutant"
139..>4153
/gene="IRS2"
/notes="N28_H29insN"
/codon_start=1
/product="insulin receptor substrate 2 insertion mutant"
/protein_id="AAK66751.1"
/db_xref="GI:14537857"
/translation="MASPPRHGPPGASGDGPNLNNNNNNNHNSVRKCGYLRKQKHG
HKRPVLRGPGAGDEATAGGSAPOPPRLEYVESEKKRSKAGAPKRVIALDCCLNI
NKRADAKHYLLALYTKDEYFAVAENEQEGHYRALTDLVSEGRAAGDAPPAAP
AASCSLIPGALGAGAGAGSISGLVAPATRAYREVMQNLKPKGLGSKNLGYI
RLCUSARTIGFKLNCQSPSVTLQLMNIRCGHSDSFFFEVGSVATSGGELMMQAD
DSVVAQNTIHETILEAMKALFELFPRPSKSSQSSGSATHPIISVFGARRHHLLVNLPP

```

```

gene
mRNA
CDS

```